UNITED STATES PATENT AND TRADEMARK OFFICE

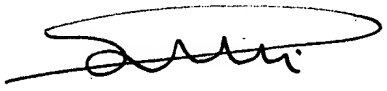
I, Béatrice Holtz, PhD,

Professional representative before the European Patent Office with Cabinet Lavoix, 2, place d'Estienne d'Orves, 75009 Paris, France declare :

1. that I am a citizen of France ;
2. that I am well acquainted with the French and English languages ;
3. that the attached is, to the best of my knowledge and belief, a true translation into the English language of the specification in French filed with the application for a patent in France on February 19, 1997, under the number FR 97 01961 ;
4. that I believe that all statements made herein of my own knowledge are true and that all statements made on information and belief are true ; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the patent application in the United States of America or any patent issuing thereon.

Béatrice Holtz

May 11, 2006



The invention relates to the use of proteins designated ULIP/POP in the diagnosis and therapy of cancers and paraneoplastic neurological syndromes.

Paraneoplastic neurological syndromes (PNS) occur in the instance of a cancer, often before its discovery, and are not connected either to the tumour proliferation itself (direct invasion, metastases) or to the therapy. Their frequency is globally estimated at approximately 1% of cancers. Several clinical pictures have been individualized for a long time (encephalomyelitis, Denny-Brown's sensitive neuropathy, cerebellar atrophy, limbic encephalitis, opsoclonus,...) corresponding in fact to the either elective or preferential attack of certain groups of neurons. The frequency of inflammatory cells in the neighbourhood of the lesions for numerous years brought to mind the possibility of an auto-immune or viral process. The more recent demonstration of auto-antibodies in the serum and the cerebrospinal fluid (CSF) of patients suffering from PNS, specific to the type of tumour and the type of neurons which degenerate, has revived the hypothesis of participation of auto-immunity in the genesis of this pathology (Graus et al., 1985; Greenlee et al., 1983).

Apart from the presence of a high titre of these antibodies in the blood and the CSF of patients, there are several arguments suggesting that PNS depend on auto-immune mechanisms. Thus the antigens recognized in the central nervous system are also present in the tumours of patients (Anderson et al., 1987). At the level of the tumour tissue, antibodies specifically directed against these antigens as well as B and T lymphocytes are found (Hetzel et al., 1990).

These data suggest that the auto-immune process could be triggered by the expression of tumour antigens. A cross-immunity process could provoke the lesions of the central nervous system. Other arguments

additionally indicate that the cerebral lesions result from the auto-immune response. Thus, in the brain of the patients, the titre of specific antibodies is higher than that of the serum and the CSF (Dalmau et al., 1991). In addition, in the case of encephalomyelitis associated with anti-Hu antibodies, there is an intense lymphocytic reaction, made up of B and T cells, situated in proximity to neurons in the process of destruction (Dalmau et al., 1991; Graus et al., 1990).

Several types of auto-antibodies allowing precise syndromic groupings as a function of immunological, neurological and carcinogenic criteria have been described.

Thus, anti-Yo antibodies are found in the serum and the CSF of women having paraneoplastic cerebellar atrophy and a gynaecological cancer (ovary, breast or uterus) (Greenlee et al., 1983; Jaeckle et al., 1985).

These antibodies recognize two cytoplasmic proteins of 34 and 62 kDa specific to Purkinje cells of the cerebellum.

The anti-Ri antibodies are found in the serum and the CSF of patients (principally of women) having opso-myoclonus, cerebellar syndrome and breast cancer. These antibodies recognize two proteins of 50 and 80 kDa specific to neurons of the central nervous system (Luque et al., 1991).

Anti-Hu antibodies are most frequently found in the course of PNS. They are found in the serum and the CSF of patients having Denny-Brown's syndrome or encephalomyeloneuritis and small-cell lung cancer (Graus et al., 1985; Dalmau et al., 1992). These auto-antibodies recognize several proteins of 37 to 45 kDa expressed specifically by all the neurons of the nervous system.

Another type of auto-antibody has recently been identified in patients having PNS: anti-CV2 antibodies (Antoine et al., 1993; Honnorat et al., 1996). The latter are atypical, in the sense that the antigenic

target recognized in adulthood is essentially non-neuronal, although the *post-mortem* analysis of the brain of four patients allows neuronal loss, gliosis and an inflammatory process characteristic of PNS to be objectivized.

The originality of the discovery of these auto-antibodies resides, on the one hand, in their demonstration. The latter escaped all the usual investigations which consisted in revealing the antigens recognized by immunohistochemistry on *post-mortem* brain. The antigen recognized is indeed soluble and disappears from *post-mortem* brain under the majority of fixation conditions. Only fixation of human *post-mortem* tissue by immersion in paraformaldehyde or *in situ* by perfusion of paraformaldehyde in animals has allowed the presence of these antibodies in the CSF or the serum of patients suffering from PNS to be revealed (Antoine et al., 1993; Honnorat et al., 1996).

The anti-CV2 auto-antibodies present in the sera of patients suffering from paraneoplastic neurological syndrome (PNS) have been defined by their capacity to recognize, by indirect immunohistochemistry, a cytoplasmic antigen expressed specifically, in adult rat brain, by a subpopulation of oligodendrocytes of the brain stem, the medulla and the cerebellum.

The originality of these auto-antibodies resides, on the other hand, in their diagnostic interest. Their presence in the serum or the CSF of patients is of diagnostic value because it allows the paraneoplastic origin of a neurological syndrome to be specified. The discovery of these antibodies, when it precedes that of cancer, directs the search to that and allows its discovery. Such was the case for six patients out of 19 having anti-CV2 antibodies. The clinical disorders were different according to the patients, certain of them having a picture of limbic encephalitis, others encephalomyeloneuritis and others Lambert-Eaton syndrome. Nevertheless, in more than 60%

of the cases, the cerebellar syndrome was predominant. The most frequently associated tumour was small-cell lung cancer (60% of the cases).

Experiments on the brains of newborn rats showed that these anti-CV2 antibodies reacted with a protein of 66 kDa (Honnorat et al., 1996).

In the adult brain, this antigen is situated in a subpopulation of oligodendrocytes or in cells which retain differentiation capacities in the adult brain (olfactory bulb, dentate gyrus). The recognized antigen could play a role in neuronal survival, via Neuron/Oligodendrocyte interactions, as the loss of neurons observed in the *post-mortem* brain of patients suffering from PNS suggests.

Its very limited expression in adulthood contrasts with a very strong and transitory expression in the central and peripheral nervous system in development, suggesting the probable role of this antigen in the development of the nervous system.

The Applicant has characterized the target antigen of anti-CV2 antibodies, which corresponds to a protein designated below by "POP-66" for "paraneoplastic oligodendrocyte protein 66 kDa".

Surprisingly, it has been discovered that the POP-66 protein belongs to the so-called ULIP family of proteins (for Unc-33-like phosphoprotein), involved in the control of neuronal development and axonal transport (T. Byk et al., 1996), and also studied in the form of CRMP proteins (Goshima et al., 1995, Wang et al., 1996), TOAD-64 (Minturn et al., 1995) and DRPs (Hamajima et al., 1996). More precisely, POP-66 has been identified as in fact being the human form of ULIP-4.

All of the data described below emphasize the complexity of this family of proteins, the existence of a very wide expression spectrum of members of this family in the brain in the course of ontogenesis, but a very limited spectrum in adults, as well as the

specificity of the anti-CV2 antibodies for a member of this ULIP protein family, which is in fact POP-66.

Thus, the Applicant has shown that the protein recognized by the anti-CV2 antibodies of patients suffering from PNS is POP-66/ULIP-4 and has established the involvement of the ULIP proteins in paraneoplastic neurological syndromes and associated cancers. In addition to their role in cancers associated with PNS, the Applicant has likewise discovered that the proteins of the ULIP family could play a role in any other form of cancer not associated with PNS. More particularly, the ULIP proteins could especially be involved in cancers of tissues having a common embryonic origin with the central nervous system.

The present invention therefore relates to a purified ULIP polypeptide, derivative or polypeptide fragment of the said purified polypeptide, comprising an amino acid sequence selected from SEQ ID No. 2, No. 4, No. 6 and No. 8.

Preferentially, the present invention relates to a purified polypeptide, derivative, or biologically active polypeptide fragment of the said purified polypeptide, comprising the amino acid sequence SEQ ID No. 8, the said polypeptide being designated by "POP-66/ULIP-4".

In the description of the invention, the following definitions are used:

- derivative: any variant polypeptide of the polypeptide of sequence SEQ No. 2, No. 4, No. 6 or No. 8 or any other molecule resulting from a modification of genetic and/or chemical nature of the sequence SEQ ID No. 2, No. 4, No. 6 or No. 8, that is to say obtained by mutation, deletion, addition, substitution and/or chemical modification of a single or of a limited number of amino acids, as well as any isoform sequence, that is to say a sequence identical to the sequence SEQ ID No. 2, No. 4, No. 6 or No. 8, to one of its fragments or modified sequences, containing one or more amino acids in the D enantiomer form, the

said modified or isoform variant sequences having conserved at least one of the properties making them biologically active.

- biologically active: having properties of induction and/or control of neuronal development and/or antigenic properties.

The invention likewise relates to an isolated nucleic acid sequence selected from SEQ ID No. 1, No. 3, No. 5 and No. 7 or a sequence derived from the sequences SEQ ID No. 1, No. 3, No. 5 and No. 7 on account of the degeneracy of the genetic code.

The various nucleotide sequences of the invention can be of artificial or non-artificial origin. They can be DNA or RNA sequences, obtained by screening of banks of sequences by means of probes elaborated on the basis of sequences selected from SEQ ID No. 2, No. 4, No. 6 and No. 8. Such banks can be prepared by conventional techniques of molecular biology known to the person skilled in the art.

The nucleotide sequences according to the invention can likewise be prepared by chemical synthesis, or alternatively by mixed methods including the chemical or enzymatic modification of sequences obtained by screening of banks.

These nucleotide sequences allow the production of nucleotide probes capable of hybridizing strongly and specifically with a nucleic acid sequence of a genomic DNA or of a messenger RNA coding for a peptide according to the invention or a biologically active fragment of this. The appropriate hybridization conditions correspond to the conditions of temperature and of ionic strength usually used by the person skilled in the art (Sambrook et al., 1989), preferably to temperature conditions of between (T_m minus 5°C) and T_m minus 30°C) and more preferably to temperature conditions of between (T_m minus 5°C) and (T_m minus 10°C) (great stringency), T_m being the theoretical melting point, defined as being the temperature at which 50% of the paired strands separate. Such probes are likewise

part of the invention. They can be used as a diagnostic tool *in vitro* for the detection, by hybridization experiments, of specific transcripts of polypeptides of the invention in biological samples or for the demonstration of aberrant syntheses or genetic anomalies resulting from polymorphism, mutations or bad splicing.

The probes of the invention contain at least 10 nucleotides, and at most contain the whole of a nucleotide sequence selected from SEQ ID No. 1, No. 3, No. 5 and No. 7 or of their complementary strand.

The *in vitro* diagnostic methods in which these nucleotide probes are employed for the detection of aberrant syntheses or genetic anomalies, such as the loss of heterozygosity and genetic rearrangement, at the level of nucleic sequences coding for a ULIP polypeptide according to the invention or a biologically active fragment are included in the present invention. Such a method type comprises:

- the contacting of a nucleotide probe of the invention with a biological sample under conditions allowing the formation of a hybridization complex between the said probe and the abovementioned nucleotide sequence, optimally after a previous amplification step of the abovementioned nucleotide sequence;
- the detection of the hybridization complex optimally formed;
- optimally the sequencing of the nucleotide sequence forming the hybridization complex with the probe of the invention.

The cDNA probes of the invention can additionally be advantageously used for the detection of chromosomal anomalies.

The nucleotide sequences according to the invention are likewise useful for the production and use of sense and/or antisense oligonucleotide primers for sequencing reactions or specific amplification reactions according to the so-called PCR technique

(polymerization chain reaction) or any other variant of this.

The nucleotide sequences according to the invention additionally have uses in the therapeutic field, for the production of antisense sequences capable of hybridizing specifically with a nucleic acid sequence, including a messenger RNA, which can be used in gene therapy. The invention thus relates to antisense sequences capable of inhibiting, at least partially, the production of a polypeptide according to the invention, such as defined above.

They are more particularly useful in the treatment of disorders of the central and peripheral nervous system and of vision, especially in the treatment of paraneoplastic neurological syndromes, as well as in anti-cancer treatment, especially of tumours associated with paraneoplastic neurological syndromes.

The nucleotide sequences according to the invention can additionally be used for the production of recombinant ULIP proteins according to the invention.

These proteins can be produced from nucleotide sequences defined above, according to techniques of production of recombinant products known to the person skilled in the art. In this case, the nucleotide sequence used is placed under the control of signals allowing its expression in a cell host.

An efficacious system of production of a recombinant protein necessitates having a vector, for example of plasmid or viral origin, and a compatible host cell.

The cell host can be selected from prokaryotic systems, such as bacteria, or eukaryotic systems, such as, for example, yeasts, insect cells, CHO (Chinese hamster ovary) cells or any other system advantageously available. A preferred cell host for the expression of proteins of the invention is formed by the bacterium *E. coli*.

The vector must contain a promoter, translation initiation and termination signals, as well as the appropriate regions of transcription regulation. It must be able to be maintained stably in the cell and can possibly possess special signals specifying the secretion of the translated protein.

These different control signals are selected as a function of the cell host used. To this end, the nucleotide sequences according to the invention can be inserted in autonomous replication vectors within the selected host, or integrative vectors of the selected host. Such vectors will be prepared according to methods currently used by the person skilled in the art, and the resulting clones can be introduced into an appropriate host by standard methods, such as, for example, electroporation.

The invention is additionally directed at the host cells transfected by these above vectors. These cells can be obtained by the introduction into host cells of a nucleotide sequence inserted into a vector such as defined above, then the culturing of the said cells under conditions allowing the replication and/or expression of the transfected nucleotide sequence.

These cells can be used in a method of production of a recombinant polypeptide according to the invention or any fragment or biologically active derivative of this.

The method of production of a polypeptide of the invention in recombinant form is itself included in the present invention, and is characterized in that the transfected cells are cultured under conditions allowing the expression of a recombinant polypeptide according to the invention or of any fragment or biologically active derivative of this, and in that the said recombinant polypeptide is recovered.

The purification processes used are known to the person skilled in the art. The recombinant polypeptide can be purified from lysates and cell extracts, from the supernatant of the culture medium,

by methods used separately or in combination, such as fractionation, chromatographic methods, immunoaffinity techniques with the aid of specific mono- or polyclonal antibodies, etc.

One variant consists in producing a recombinant polypeptide fused to a "carrier" protein (chimeric protein). The advantage of this system is that it allows a stabilization and a decrease in the proteolysis of the recombinant product, an increase in the solubility in the course of the *in vitro* renaturation and/or a simplification of the purification when the fusion component has an affinity for a specific ligand.

The exploitation of ULIP proteins, and in particular POP-66/ULIP-4, as well as antibodies directed against these proteins, is promising in various fields.

Thus, the detection of the anti-CV2 auto-antibody by immunofluorescence on fixed animal brain is currently used as a diagnostic test.

The production of POP-66/ULIP-4 recombinant protein according to the invention allows the production of a rapid and reliable test (of Elisa or Western Blot type) for detecting anti-CV2 antibodies.

Such tests already exist for anti-Hu, anti-Yo and anti-Ri antibodies. The test for detecting anti-CV2 in the serum of patients could be prescribed in the case of suspicion of paraneoplastic neurological syndrome and consequently could include anti-CV2 antibodies at the same titre as the other antibodies identified in the PNS such as mentioned above.

The invention is therefore likewise directed at a method for the diagnosis of paraneoplastic neurological syndromes and/or for the early diagnosis of the formation of tumours of cancerous origin, characterized in that auto-antibodies directed against a POP-66/ULIP-4 protein are demonstrated in a blood sample taken from an individual by

- the contacting of a blood sample taken from an individual with a purified polypeptide (POP-66), derivative or biologically active polypeptide fragment of POP-66/ULIP-4 optionally attached to a support under conditions allowing the formation of specific immunological complexes between the said polypeptide and the auto-antibodies optionally present in the serum sample, and

- the detection of the specific immunological complexes optionally formed.

The invention likewise relates to a kit for the diagnosis of paraneoplastic neurological syndromes and for the early diagnosis of the formation of tumours from a biological sample, comprising:

- at least one purified POP-66/ULIP-4 polypeptide, derivative or biologically active polypeptide fragment of POP-66/ULIP-4, optionally attached to a support,

- means of visualization of the formation of specific antigen/antibody complexes between an anti-POP-66 auto-antibody and the said purified POP-66 polypeptide, derivative or polypeptide fragment and/or means of quantification of these complexes.

The invention likewise relates to mono- or polyclonal antibodies or their fragments, chimeric or immunoconjugated antibodies obtained from a purified ULIP polypeptide comprising an amino acid sequence selected from SEQ ID No. 2, No. 4, No. 6 and No. 8, derivative or biologically active polypeptide fragment of ULIP and their use for the purification or the detection of a ULIP protein in a biological sample.

Polyclonal antibodies can be obtained from the serum of an animal immunized against the protein, produced, for example, by genetic recombination according to the method described above, according to the usual working methods.

The monoclonal antibodies can be obtained according to the conventional method of hybridoma culture described by Köhler and Milstein.

The antibodies can be chimeric antibodies, humanized antibodies, Fab and $F(ab')_2$ fragments. They can likewise be present in the form of immunoconjugates or labelled antibodies.

The invention likewise relates to the use of antibodies directed against a protein of the ULIP family for the demonstration of a ULIP protein in neoplasms, and paraneoplastic neurological syndromes for diagnostic purposes.

Preferentially, the invention relates to the use of monoclonal antibodies obtained from polyclonal anti-CV2 serum of patients by immortalization of lymphocytes, according to the usual techniques known to the person skilled in the art.

Thus, the antibodies directed against a protein of the ULIP family are useful for detecting abnormal expression of ULIP protein in patients having neurological syndromes, in whom cancer has not been diagnosed by the conventional methods. This abnormal expression of ULIP protein will be able to be correlated with the existence of a cancer which had not been spotted. Thus, the antibodies directed against a ULIP protein, especially against POP-66/ULIP-4, are useful for the early diagnosis of cancer.

The invention likewise relates to a method of determination of an allelic variability, a mutation, a deletion, an insertion, a loss of heterozygosity or a genetic anomaly of the POP-66/ULIP-4 gene, situated on chromosome 10 in the 26q region and which can be involved in pathologies, characterized in that it employs at least one nucleotide sequence SEQ ID No. 7. Amongst the methods of determination of an allelic variability, a mutation, a deletion, an insertion, a loss of heterozygosity or a genetic anomaly of the POP-66/ULIP-4 gene, a method comprising at least one PCR amplification step of the nucleic sequence of POP-66/ULIP-4 capable of having a polymorphism, a mutation, a deletion or an insertion with the aid of pairs of primers of nucleotide sequences, a step in the

course of which amplified products are treated with the aid of appropriate restriction enzymes and a step in the course of which at least one of the products of the enzymatic reaction is detected or determined is preferred.

Advantageously, it is possible to search for the mutations associated with the said chromosome 10 in relation to cancer, especially peripheral cancerous tumours and primitive cerebral tumours of glial origin, for example.

The invention likewise relates to a pharmaceutical composition comprising at least one purified protein of the ULIP family, polypeptide fragment or biologically active derivative of this, a nucleotide sequence or nucleotide sequence fragment coding for the said protein, an antisense sequence capable of hybridizing specifically with a nucleotide sequence coding for the said protein, or an antibody directed against the said protein, combined with a pharmaceutically acceptable vehicle.

The invention preferentially comprises pharmaceutical compositions comprising as active principle a purified POP-66 polypeptide, derivative or polypeptide fragment of POP-66, preferentially in soluble form, combined with a pharmaceutically acceptable vehicle.

Such compositions offer a new approach to treating disorders of the central and peripheral nervous system and of vision, and especially paraneoplastic neurological syndromes. In addition, they are useful for treating neurological disorders connected with a neuronal loss and/or an underexpression of ULIP proteins in the nervous system.

Thus, POP-66/ULIP-4 is also of interest in neurodegenerative pathologies such as multisystemic atrophies which are conditions similar to those of PNS and for which an anomaly of an oligodendrocytic subpopulation has been detected (Papp et al., 1992).

The compositions according to the invention are additionally useful in anticancer therapy.

The antibodies directed against one or more ULIP proteins can be combined with antineoplastic agents, thus allowing the targeting of medicaments towards the tumour cells.

They can additionally be combined with a hydrophilic chemical group chosen in such a way so as to cross or not to cross the blood-brain barrier, according to the type of tumour.

The ULIP proteins and in particular POP-66 as well as the nucleotide sequences coding for the said proteins and the antisense sequences or oligonucleotides can be useful in the therapy of any type of cancer in which a gene coding for a ULIP protein is involved. Amongst examples of cancers, it is possible to mention peripheral tumours, such as small-cell lung cancer, thymoma, cancer of the breast and of the ovary, as well as cerebral tumours, preferably primitive cerebral tumours of glial origin. The expression of POP-66 in the non-proliferative cells of normal brain, its absence in normal tissues such as lung or thymus, for example, its differential reexpression during tumorigenesis of these tissues and the modulation of its expression in a tumour line in the course of differentiation suggest in this respect that POP-66 could be a tumour suppressor gene.

Preferentially, the pharmaceutical compositions according to the invention can be administered by the systemic route, preferably by the intravenous route, by the intramuscular route, intradermally or by the oral route.

Their modes of administration, dosages and optimal pharmaceutical forms can be determined according to the criteria generally taken into account in the establishment of a therapeutic treatment adapted to a patient, such as, for example, the age or the body weight of the patient, the seriousness of his/her

general condition, the tolerance to the treatment and the secondary effects noted, etc.

The invention likewise comprises the use of a purified protein of the ULIP family, polypeptide fragment or biologically active derivative of this, a nucleotide sequence or nucleotide sequence fragment coding for the said protein, an antisense sequence capable of hybridizing specifically with a nucleotide sequence coding for the said protein, or an antibody directed against the said protein, combined with a pharmaceutically acceptable vehicle, for the production of a medicament intended for treating neurodegenerative illnesses and neoplasms.

The examples and the figures whose legends are presented below are given by way of illustration.

LEGEND TO THE FIGURES

- Figure 1 represents a two-dimensional electrophoresis profile obtained from brain protein extracts of newborn rats enriched in POP-66.

A: silver staining of all of the proteins.

B: immunoblot with the anti-CV2 serum of patients.

The arrows indicate the spots corresponding to POP-66, revealed with anti-CV2 antibodies.

- Figure 2 represents a two-dimensional electrophoresis profile obtained from protein extracts of brains of newborn rats.

Immunoblot with A- antipeptide antibody 3 and B- anti-CV2 antibody.

- Figure 3 represents a one-dimensional electrophoresis obtained from protein extracts of brains of newborn rats.

Immunoblot with a: preimmune serum for peptide 3

Immunoblot with b: anti-peptide serum 3

Immunoblot with c: anti-peptide serum 4

Immunoblot with d: preimmune serum for peptide 4.

- Figure 4 represents an immunohistochemical labelling of sections of brains of adult rats with

A: anti-CV2 serum of a patient suffering from PNS

B: rabbit serum with anti-peptide 3 antibodies

C: rabbit serum with anti-peptide 4 antibodies.

- Figure 5 represents a histological labelling of sections of young rat cerebellum 8 days post-natally.

A: Staining with toluidine blue; ge = external granular layer; m = molecular layer (×400).

B: Immunolabelling after incorporation of BrdU (bromodeoxyuridine). The cells which have incorporated BrdU are virtually all situated in the most external zone of the external granular layer (ge). Some positive cells are situated in the internal granular layer (×400).

C: Indirect immunoperoxidase with a patient serum containing an anti-CV2 antibody (×400). The immunoreactivity is concentrated in the internal part of the external granular layer (future molecular layer (m)). Some cells are immunoreactive in the internal granular layer. The Purkinje cells (p) are negative as well as the cells of the external part of the external granular layer (ge).

D: Indirect immunoperoxidase with a patient serum containing an anti-CV2 antibody (×1000). Above all, the immunolabelling is concentrated in the internal part of the external granular layer (future molecular layer (m)). A reactive cell is noted in the internal granular layer (gi) (arrow).

- Figure 6 represents immunohistochemical labelling of sections of post-mortem hippocampus (HPS staining).

A: brain of control patient,

B: brain of patient having limbic encephalitis, and circulating anti-CV2 antibody. It is possible to note the disappearance of the granular cells.

- Figure 7 represents a two-dimensional electrophoresis profile with the control ULIP-2 protein (A) and the ULIP-4 protein (B).

Figure 7C represents the migration profile model of the proteins ULIP-1, 2, 3 and 4 as a reference.

The proteins are revealed:

a) by autoradiography to locate the proteins translated *in vitro* (translation);

b) by immunoblotting with the anti-CV2 serum.

- Figure 8 represents a migration profile of the mRNA of C-22/ULIP-3 (8A) and TOAD-64/ULIP-2 (8B) amplified by RT-PCR expressed in different cell types:

lanes 1-3: small-cell lung tumour

lane 2: small-cell lung tumour with anti-CV2 serum

lane 4: control cDNA.

lane 5: medulloblastoma treated by HTLV1 infection

lanes 6-7: medulloblastoma

lane 8: C6 line of glial cells in mice

lane 9: control

lane 10: nothing

lane 11: kb scale.

The black arrows correspond to POP-66; the white arrows correspond to the molecular weight standard.

- Figure 9 represents the nucleotide sequence of ULIP-2 in mice (SEQ ID No. 1), as well as the inferred amino acid sequence (SEQ ID No. 2).

- Figure 10 represents the nucleotide sequence of ULIP-3 in mice (SEQ ID No. 3), as well as the inferred amino acid sequence (SEQ ID No. 4).

- Figure 11 represents the nucleotide sequence of ULIP-4 in mice (SEQ ID No. 5), as well as the inferred amino acid sequence (SEQ ID No. 6).

- Figure 12 represents the nucleotide sequence of ULIP-4 in man (SEQ ID No. 7), as well as the inferred amino acid sequence (SEQ ID No. 8).

An erroneous stop codon in the human ULIP-4 sequence (asterisk) arises from a fault of the reverse transcriptase in the production of the bank. By comparing with ULIP-4 of rats and of mice, it is almost certain that the TAG sequence coding for a stop is in fact an AAG codon, coding for a lysine as in rats and mice. In addition, the region around this amino acid is entirely conserved in the three species.

The amino acid sequence has been completed in SEQ ID No. 8 by 15 C-terminal amino acids (No. 554 to No. 568). This C-terminal region which is missing in Figure 12 is very well conserved between rat and mice ULIP-4 as well as between the different ULIPs.

EXAMPLE 1:

Purification of POP-66 and sequencing

The purification of POP-66 is carried out according to the material and the methods described in the article of Honnorat et al., 1996, incorporated by reference, starting from serum of patients suffering from PNS.

To identify the protein POP-66, a purification strategy was chosen which allows a partial sequencing to be obtained. The screening of an expression bank of brain cDNA or the immunoaffinity purification of the protein were excluded because of the limited quantities of sera linked to the death of the patients. It was possible to develop a method of biochemical purification starting from brains of newborn rats on account of the anti-CV2 human sera, which allowed each purification step to be monitored.

The tissues, preserved at -70°C before use, were treated with a solution containing 0.2 M DTT (dithiothreitol) (Sigma) 2% Ampholine 3-10 (Pharmacia), 2% Triton X-100 (Merck) and placed at 2-4°C. Immediately before use, solid urea (Pharmacia) was added to obtain an 8M solution.

The POP-66 protein is soluble, at least in part, and precipitates entirely at a concentration of 40% ammonium sulphate.

Centrifugation at 100,000 (times) g and ammonium sulphate precipitation (eliminating the proteins precipitating below 20% and above 40% ammonium sulphate) allows protein extracts enriched in POP-66 to be obtained. The proteins of this extract are then separated, after dialysis, by isofocussing on agarose gel (Peltre et al., 1982).

After transfer to a membrane, the anti-CV2 antibodies recognize several bands of isoelectric points of between 5.85 and 6.55. All of these bands correspond to the POP-66 protein recognized by the anti-CV2 antibodies. This spectrum suggests the possibility of transcriptional modifications (phosphorylations and/or glycosylations) of the protein.

The zone of proteins of pI between 5.85 and 6.55 from the agarose gel is used for à new electrophoretic migration in denaturating medium on polyacrylamide gel previously equilibrated with an equilibration solution (0.05 mol/l Tris/HCl, pH 6.8, 6M urea, 30% glycerol, 1% weight/volume SDS for 2 x 10 minutes) to which is added DTT (0.25% weight/volume) and bromophenol blue.

Two methods of detection are used:

- *silver staining*. Immediately after the end of the migration, the gel is immersed in a fixing solution (40% ethanol; 10% acetic acid) for 30 minutes; it is then placed in an incubation solution (30% ethanol, 7% weight/volume of sodium acetate, 0.1% glutaraldehyde, 0.2% weight/volume of sodium thiosulphate) for 30 minutes or one night. After washing, the gel is placed in a silver solution (0.1% weight/volume of silver nitrate + formaldehyde) and developed (2.5% weight/volume of sodium carbonate + formaldehyde). The reaction is stopped with Na_2 EDTA (1.5% weight/volume). The gels are preserved in a glycerol solution.

- *transfer to a PVDF membrane* (Immobilon-P®, Millipore). The separated proteins are transferred to a PVDF membrane using a 100 mM CAPS buffer (Sigma) of pH 11. The transfers are incubated for one hour in TBS buffer (Tris buffer saline) with 5% of casein (milk) and 18 hours in TBS buffer (+ 1% of casein) containing antibody (1/500 anti-CV2 serum). After washing with TBS-casein (15 minutes), visualization is carried out by incubating the transfers for 1 and a half hours with biotinylated anti-IgG antibodies (1/1000) and for 1 and a half hours with the streptavidin-peroxidase complex (1/2000). The transfers are then visualized with DAB (0.06% weight/volume diaminobenzidine in 0.05 M Tris) and with H_2O_2 (0.02 µg/ml).

A single band corresponding to a protein of 66 kDa is visible. This is specifically labelled with anti-CV2 antibodies (Figure 1). An N-terminal sequencing of this protein was then carried out, after trypsic digestion.

Seven peptides, having the following sequences, were obtained:

1 - X-Met-Tyr-Asp-Gly-Pro

2 - X-Phe-Asn-Leu-Tyr-Pro-Arg

3 - X-Val-Leu-Glu-Asp-Gly-Thr-Leu-His-Val-Thr-Glu-Gly

4 - X-Ile-Gly-X-X-Ala-Gln-Val-(His ?)-Ala-Glu-Asn-Gly-X-Ile-Ile-Ala-Glu-Glu-Gln

5 - X-X-Glu-Asn-Gln-Phe-Val-Ala-Val-Thr

6 - X-Val-Asn-Asp-(Asp ?)-Gln-Ser-Phe-Tyr-Ala-Asp-Ile-Tyr-Met-Glu-(Asp ?)-(Gly ?)-Leu-Ile

7 - X-X-X-Phe-Val-Thr-X-Pro-X-Leu-X-Pro

X: corresponds to a non-determined amino acid,

(?): corresponds to a probable but uncertain amino acid.

According to the analysis of databanks available in 1994, no known protein corresponded to these sequences.

EXAMPLE 2

Cloning of the cDNA of POP-66 or of related proteins

The cloning of the cDNA of the POP-66 protein or of related proteins was carried out by using degenerate oligonucleotide probes obtained from fragments of two peptides:

Ile-Ile-Ala-Glu-Glu-Gln

Tyr-Ala-Asp-Ile-Tyr-Met-Glu-(Asp ?)

Four sets of degenerate oligonucleotide primers (sense/antisense) are therefore determined (AT(C/T)ATTGC(T/A)GA(A/G)CA;TG(C/T)TC(T/C)AC(T/A)-GCAT(A/G)AT; TATGC(A/T)GA(C/T)AT(C/T)ATGGA; TCCAT(G/A)TA(G/A)CT-(T/A)GCATA, and used for a PCR amplification.

The matrix is prepared in the form of double-stranded cDNA (Promega kit) from poly(A^+)RNA extracted from the brain of rats 10 days old (Zivic-Miller, USA) using the Fast Track mRNA isolation kit (Invitrogen).

The conditions of amplification by PCR are as follows: 35 cycles at 94°C, 1 minute for denaturation, 55°C, 1 minute for hybridization and 72°C, two minutes, for extension.

The PCR products are analysed by 1% agarose gel electrophoresis, electroeluted, cloned in a TA cloning vector (Invitrogen) and sequenced using the primer sites of the T7 and SP6 promoters.

The sequence of amino acids inferred from the MFB-17 clone agrees with the sequences of the two original peptides of POP-66 determined by the analysis of the amino acid sequence.

A comparative analysis of the nucleic acid sequences using the Genbank and EMBL databases reveals that MFB-17 is a partial cDNA with a nucleotide sequence identical to that of a segment of TOAD-64, a rat neuronal protein (Minturn et al., 1995).

The amino acid sequence inferred in the cDNA of TOAD-64 agrees with the sequences of the seven peptides determined by partial sequence analysis of the protein recognized by the anti-CV2 antibodies after purification by electrophoresis.

The molecular weight, the isoelectric point, the immunohistochemical profile and the regulation of TOAD-64 are similar to those of the POP-66 antigen.

Since the MFB-17 clone did not have the complete coding region, it was necessary to produce an intact recombinant protein to continue the research concerning the CV2 protein.

To obtain a complete TOAD-64 protein, the ds-cDNA matrix of rat brains was amplified with two sets of primers situated at the 5' and 3' extremities of the coding regions
(sense: GGCATATGTCTTATCAGGGGAAG;
antisense: GCGAATTCTTAGCCCAGGCTGATG).

This approach allowed two different clones to be produced, one corresponding to the TOAD-64 sequence and the other to a clone designated by C-22.

EXAMPLE 3:

Comparison of the amino acid sequence inferred from C-22 with the ULIP proteins

The amino acid sequence inferred from the open reading frame indicates that this C-22 clone belongs to the superfamily of ULIP genes represented by several genes of EST sequences.

The amino acid sequence inferred from C-22 has a homology of 30% with the amino acid sequence of the unc-33 protein of *Caenorhabditis elegans*.

Recently, four different homologous genes in the unc-33 protein have been described in mammals and the chicken.

An analysis of the sequences by the Genbank databases and protein banks has allowed a classification of the unc-33- like (ULIP) proteins into four different subgroups to be proposed (Byk et al. 1996).

However, as the real functions of these proteins are not clearly known, the proposed classification is simply based on the percentage of identity of amino acids. ULIP-1 is represented by a mouse "unc-33-like" phosphoprotein which has a homology

of 76% with TOAD-64, Crmp-62 and Munc, a mouse sequence recently available from Genbank.

ULIP-2 is composed of TOAD-64, Crmp-62 and Munc which between them have a 97% amino acid identity.

The partial human EST sequences, that is to say hcrmp-1, which have a 75% identity with ULIP-1 or ULIP-2, have been found. They belong to a third group called ULIP-3. The last group identified called ULIP-4 comprises r-CRMP-3 in the rat and the forms ULIP-4 in the mouse and POP-66/ULIP-4 in man.

The comparison of the amino acid sequence of the three ULIP genes, namely TOAD-64 in the rat, Crmp-62 in the chicken and ULIP-1 in the mouse, with the amino acid sequence inferred from the open reading frame of the present C-22 clone, using the Clustal V alignment software program reveals that C-22 has an identity of 74% with ULIP-1, 77% with Crmp-62 and 76% with TOAD-64.

The nucleotide sequence C-22 has an identity of 97% with the partial sequence EST, hCrmp-1, and thus defines the third member of the ULIP-3 group. The TOAD-64, Crmp-62 and C-22 genes each code for a protein of 572 amino acids in length, whereas the amino acid sequence inferred from ULIP-1 gives a protein of 570 amino acids.

The analysis of the amino acid sequence of C-22 does not show any signal sequence or transmembrane domain suggesting that the product(s) of the C-22 gene could be localized in the cytoplasm of the cells.

Several consensus sites of phosphorylation by the kinase C protein (S/T X R/K) appear along the length of the product of the C-22 gene. These observations suggest that C-22 is a phosphoprotein and that the slight differences in the phosphorylation could dictate the activity or the role of different members of this family of proteins throughout the cell cycle.

Table I: Summary of proteins having a homology with the ULIPs.

Family		Species	EMBL No.
Nematode Unc-33		Nematode	Z14146
Dihydropyrimidinase	Hu DHPase	human	D78011
	Ra DHPase	rat	D63704
ULIP-1 group	Ulip	mouse	X87817
	Hu DRP3	human	D78014
	r-CRMP-1	rat	U52102
	Hu-Ulip	human	Y07818
ULIP-2 group	ULIP-2	mouse	SEQ ID No. 2
	Toad-64	rat	Z46882
	CRMP-62	chicken	U17277
	Munc	mouse	X87242
	HCRMP-2	human	U17279
	Hu DRP-2	human	D78013
	r-CRMP-4	rat	U52104
ULIP-3 group	ULIP-3	mouse	SEQ ID No. 4
	HCRMP-1	human	U17278
	rCRMP-1	rat	U52102
	C-22	rat	U52095
	Hu DRP-1	human	D78012
ULIP-4 group	ULIP-4	mouse	SEQ ID No. 6
	POP-66/ULIP-4	man	SEQ ID No. 8
	r-CRMP-3	rat	U52103

EXAMPLE 4:

Regulation of the expression of the C-22 gene:

The evaluation of alterations in the expression of the C-22 gene could have considerable significance for the knowledge of the functional aspects of the C-22 protein.

Consequently, the Applicant studied the possible regulation of the expression of the C-22 gene

in the course of development. The total RNA is extracted and separated by electrophoresis on 1% agarose gel and transferred to Nytran membrane (Duchemin et al. 1987). The transfers are hybridized with a C-22 coding sequence labelled with ^{32}P, a 0.5 mM phosphate buffer and 5% SDS at 65°C for 16 hours.

At the end of the hybridization, the transfers are washed successively three times with 2 × SSC, 0.1% SDS at ambient temperature, then 1 × SSC, 0.1% SDS at 65°C for 60 minutes, and exposed to X-rays.

Under the conditions used, a single band at 3.8 kb was detected representing the C-22 mRNA which is also the smallest transcript of the unc-33 family of genes of vertebrates. The size of the transcript remains the same during the pre- and post-natal periods.

The kinetics of the C-22 gene in the brain of rats in the course of development shows that the messenger is detectable in the course of the embryonic period on day E17. The quantity of C-22 transcripts increases up to day 7 post-natally then decreases rapidly from the second week after birth to a virtually undetectable level in the adult.

Around birth, a still unknown regulation signal is probably received, which increases the expression of the C-22 gene, this signal being temporarily linked to neuronal differentiation and to axonal development.

The mRNA of C-22 has not been able to be detected by Northern Blot analysis in several regions of the brain such as the frontal cortex, the midbrain and the thalamus in adults and rats more than two years old.

In addition, it has not been possible to detect the mRNA of C-22 in non-neuronal tissues, such as the heart, the lung, the liver and the kidney in one-week old rats and adult rats.

The data on the expression profile of the mRNA of C-22 suggests a decisive role of the C-22 protein in the development of the brain.

EXAMPLE 5:

Immunoblotting of POP-66:

Antibodies against four of the sequenced peptides were products (see example 1). Two of the sera turned out to be of particular interest.

One contains antibodies (Ab anti-pep3) which recognize several members of the ULIP family on two-dimensional electrophoresis of protein extracts of newborn rat brain (Figure 2) and on one-dimensional electrophoresis (Figure 3). On Western Blotting, another antibody (Ab anti-Pep4) recognizes a single band of 66 kDA capable of corresponding to a single member of the family (Figure 3).

EXAMPLE 6:

Immunohistochemistry

The tissue preparations for immunohistochemistry are obtained from newborn rat brains and from *post-mortem* human brains, fixed at 4°C in 4% paraformaldehyde and 0.2% picric acid diluted in phosphate buffer (0.1 M, pH = 7.4), then cryoprotected.

Immunocytochemistry can be carried out by the indirect immunofluorescence technique. Sections of 12 µm in thickness are prepared in a cryostat and then mounted on gelatin-covered slides, treated with 0.1% Triton X100 for 2 hours in PBS buffer and 1% bovine serum albumin (BSA) and incubated for 12 h with anti-CV2 serum of patients in PBS/1% BSA at ambient temperature (1/100 dilution of the serum). After several washes with PBS/1% BSA, the sections are incubated for 2 h with a rabbit anti-human antiserum conjugated to fluoroscein, diluted to 1% (Dakopatts) in PSB/1% BSA. After washing in PBS, the slides are examined under the microscope. The control sections are incubated either with anti-human IgG antiserum conjugated to fluoroscein alone, or PBS/1% BSA alone, or the patient serum alone, or finally the control serum (patients not suffering from PNS) and antibodies conjugated to fluoroescein at the same dilution.

To confirm the immunofluorescence, it is possible to use indirect labelling by immunoperoxidase. The frozen tissue sections fixed with paraformaldehyde are incubated with 0.3% H_2O_2 (to destroy the intrinsic peroxidase activity) and 10% normal rabbit serum (to avoid the non-specific binding of the rabbit IgG) or 1% BSA. After incubation for 12 h with patient sera diluted to 1/1000 and washing, the sections are incubated for 2 h with biotinylated rabbit anti-human IgG antiserum diluted to 1/1000 in PBS/1% BSA. The bound human IgGs are visualized by incubation with an avidin-biotin-peroxidase complex (Vectastain ABC complex, Vector) and developed with 0.05% DAB (Sigma). The control sections are obtained with sera of 15 patients without PNS according to the same protocol.

A - Localization of proteins of the ULIP family with the aid of antipeptide antibodies:

Immunohistochemical labelling was carried out on sections of newborn and adult rat brains. The antipeptide-3 antibody recognizes (an) antigen(s) present in several cell types on sections of newborn and adult rat brains (Fig. 4). Like the patient anti-CV2 serum, the anti-peptide-4 antibodies do not allow the demonstration of any antigen on sections of newborn rat brain although they specifically label a subpopulation of oligodendrocytes in adult rat brain (Fig. 4).

B - Expression of POP-66 in the course of the normal development of the brain:

Figure 5 shows that the proliferative nerve cells of the progenitor zones of the nervous system demonstrated by the accumulation of bromodeoxyuridine (BrdU) do not express POP-66 although the non-proliferative cells which correspond to the nerve cells in differentiation or in migration express it.

EXAMPLE 7:

Role of POP-66 in neuronal survival

Figure 6 allows human brain sections of healthy patients and of patients suffering from PNS to be compared. In the patients suffering from PNS and having circulating anti-CV2 antibodies, a disappearance of the neurons of the dentate gyrus and of pyramidal neurons (central cell band), as well as an intense astrocytic reaction, are observed.

EXAMPLE 8:

Characterization of the POP-66 protein - identification with ULIP-4:

Materials and methods

a) Partial purification of ULIP-1

Partially purified ULIP-1 was obtained from newborn mouse brains by three purification steps. These brains were homogenized in 4 volumes of homogenization buffer (25 mM sodium phosphate, pH 7.8, 1 mM EGTA, 10 µg/ml of leupeptin, 25 µg/ml of aprotinin, and 10 µg/ml of pepstatin. The homogenates were centrifuged for 10 minutes at 400 × g. The plugs were resuspended in 2 volumes of homogenization buffer, homogenized and centrifuged again. The supernatants from two centrifugations were collected, sonicated and centrifuged for 1 hour at 100,000 × g. The supernatant (S2) was loaded onto a column of DEAE-Sepharose CL-6B (1.75 cm^2 × 26 cm) equilibrated with 100 ml of buffer A (25 mM sodium phosphate, pH 7.8, 1 mM EGTA) at a flow of 30 ml per hour. The proteins were eluted in 300 ml of a 0-250 mM linear gradient of sodium chloride in buffer A and 5 ml samples were collected. The fractions containing ULIP were collected and solid ammonium sulphate was added to 20% saturation. This pool was loaded onto a column of phenyl-Sepharose CL-4B (1.75 cm^2 × 22 cm) which had been previously equilibrated with 100 ml of buffer B (10 mM sodium phosphate, pH 7.8, 1 mM EGTA) containing 20% of saturated ammonium sulphate. The proteins were eluted

in a linear gradient decreasing from 20 to 0% of saturated ammonium sulphate in buffer B. The fractions containing ULIP were collected and dialysed twice against 20 volumes of buffer A. The proteins were concentrated in a small (10 ml) column of DEAE-Sepharose CL-6B and eluted with 400 mM sodium chloride in buffer A. The eluate was desalted on a Sephadex G-25 (NAP-10) column and concentrated to a final volume of 0.5 ml by evaporation. In the last purification step, the concentrated fraction was chromatographed in three successive steps, on two Superose 12 FPLC (Fast Protein Liquid Chromatography) columns mounted in series, in buffer C (50 mM sodium phosphate, pH 7.2, 150 mM sodium chloride) at a rate of 0.3 ml/minute. The fractions (0.6 ml) were collected and the fractions enriched in ULIP were analysed. The presence of ULIP in the successive purification steps was tested by a one-dimensional Western Blot using an anti-stathmin antibody capable of cross-reactivity. The proteins were quantified according to the method of Bradford.

b) Migration on electrophoresis gel:

A one-dimensional electrophoresis was carried out on 13% polyacrylamide gels according to the method of Laemmli. The two-dimensional PAGE electrophoreses were carried out as described above. The isoelectro-focussing gels contained 2% of total ampholines, pH 6-8 and 3-10 in a ratio of 4:1. The second dimension had been carried out on 10% acrylamide gels. The proteins had been either subjected to immunoblotting or stained with silver.

c) Western Blot analysis:

The proteins were transferred from gels to nitrocellulose in buffer containing 48 mM Tris, 39 mM glycine and 5% of methanol. The membrane was saturated with casein (2.5%) in the immunoblotting solution (12 mM Tris-HCl, pH 7.4, 160 mM NaCl, 0.1% Triton X-100) and tested with an antiserum directed against the peptide I of rat stathmin (1/10,000 dilution) or an antiserum directed against the recombinant ULIP protein

(dilution 1/20,000) diluted in an immunoblotting solution containing 1% of casein. The bound antibodies were detected either with a protein A labelled with ^{125}I and autoradiographed or with anti-rabbit antibodies bound to peroxidase using the ECL kit (Amersham).

d) Analysis of the protein sequence:

The fractions enriched in ULIP were separated on polyacrylamide gels in two dimensions. The gels are fixed for 30 minutes in 25% ethanol and 10% acetic acid and stained for 3 minutes in 0.1% amido black in 1% acetic acid and 40% methanol. The gels were decolourized in 1% acetic acid and the spots corresponding to the principal form of ULIP were cut out in these three gels, collected and digested with 2 mg/ml of Lys C endoprotease. The peptides eluted from the gel were then separated by HPLC on a DEAE-C18 column with a gradient of 0-55% of acetonitrile in 0.1% trifluoro-acetic acid. The peptides were then sequenced according to the Edman automatic degradation.

e) Expression *in vitro* in a mammal

1 µg of the Bluescript plasmid containing the entire cDNA coding for ULIP-1, ULIP-2, ULIP-3 or ULIP-4 was used to carry out the transcription and translation *in vitro* with the "Reticulocyte lysate" system (Promega) according to the protocol described by the manufacturer. 5 µg of the total transcription/translation mixture of 25 µl were analysed on electrophoresis gel in two dimensions.

Results

Neither the recombinant protein ULIP-1, nor the recombinant proteins TOAD-64 (ULIP-2) and C-22 (ULIP-3) were recognized by the anti-CV2 sera. In addition, the distribution profile of the spots corresponding to POP-66 recognized by the anti-CV2 antibodies on electrophoresis in two dimensions does not correspond to the spots recognized by the anti-ULIP-1 antibodies. However, POP-66 is a member of the ULIP family since the three POP-66 spots are recognized by the anti-pep3

Ab. POP-66 therefore corresponds to a member of the family of more basic pH_i.

After translation *in vitro* of the four proteins (ULIP-1, 2, 3 and 4), it was shown that ULIP-4 has the same 2D electrophoretic profile as POP-66 and is recognized by the anti-CV2 antibodies (Figure 7).

For this, the ULIP-4 protein and, as control, the ULIP-2 protein were translated *in vitro* in the presence of ^{35}S methionine from cDNA clones coding for the entire proteins. The proteins were separated by two-dimensional electrophoresis (in the presence of a brain extract providing the essential references), transferred to nitrocellulose and visualized:

- by autoradiography to localize the proteins translated *in vitro* (translation);
- by immunoblotting with the CV2 serum.

Figure 7 shows that the three spots from the *in vitro* translation of ULIP-4 correspond to the spots recognized by CV2. These spots are not recognized in the translation of ULIP-2.

The CV2 serum therefore specifically recognizes ULIP-4.

This allowed POP-66 to be identified like ULIP-4.

EXAMPLE 9:

Chromosomal localization of the POP-66/ULIP-4 protein

Having cloned the cDNA of human ULIP-4, it is then possible to determine the chromosomal localization of the POP-66/ULIP-4 gene by genetic mapping by *in situ* isotopic hybridization (Levy and Mattei et al., 1995).

In situ hybridization is carried out on preparations of chromosomes obtained from human lymphocytes stimulated by phytohaemagglutinin cultured for 72 hours. 5-Bromodeoxyuridine was added during the last 7 hours of culture (60 µg/ml of medium) to ensure a post-hybridization image of chromosome bands of good quality. The clone containing an insert of 1300 base

pairs coding for ULIP-4 in the Bluescript vector is labelled with tritium by nick translation with a specific activity of 1×10^8 dpm. μg^{-1}. The radiolabelled probe was hybridized in the metaphase stage at a final concentration of 200 ng per ml of hybridization solution. After covering with a Kodak NTB_2 emulsion, the slides were exposed for 20 days at +4°C and then developed. To avoid the shifting of the silver grains during the process, the chromosome blots were previously labelled with a Giemsa buffer solution and the metaphases were photographed. The visualization of the bands was carried out by the "Giemsa fluorochrome photolysis" (FPG) method and the metaphases were rephotographed before analysis. Of the 100 cells in metaphase examined after hybridization *in situ*, 246 silver grains associated with the chromosomes were counted and 54 amongst these (21.9%) were localized on chromosome 10. The distribution of the grains on this chromosome was not random: 39 out of 54 (72.2% of the latter) were localized on the q25.2-q26 region of the long arm of chromosome 10.

The POP-66/ULIP-4 gene is therefore found to be situated on chromosome 10 in the q25.2-q26 region. This region has been described to be deleted in some cases of glioblastome. This chromosomic localisation confirms the interest in studying POP-66/ULIP-4 in connection with tumoral pathology.

EXAMPLE 10

Expression of ULIP proteins in transfected HeLa cells

A- Materials and methods : RT-PCR experiments :

Total RNA was extracted using 1 mL RNAZOL^{TM}B (Bioprobe) according to the method of Chomczynski and Sacchi. RNA quantity was determined by optical density measured at 260 nm and its purity was determined by the ratio of absorbances measured at 260 and 280 nm (ratios 1.8-2.0).The integrity of RNA preparations was further

checked by electrophoresis on 1% agarose gel in TBE (Tris-borate 0.45 M, EDTA 10 mM, pH 8).

Primer specificity was analyzed by comparing these sequences with various gene databanks (EMBL and FASTA). For relative quantitation, G3PDH encoding gene (glyceraldehyde-3-phosphate deshydrogenase, Clontech), an ubiquitous gene expressed in many tissues including brain, was co-amplified with the assayed mRNA as an internal control in order to check efficiency of the reverse transcription step for the different RNA samples. The 5', 3' primers and the internal probe oligonucleotides for G3PDH were synthesized and purified by Eurogentec. Total mRNA (1 μg) was denatured (15 min at 65°C) and transcribed into single strand cDNA (1 hour and half, 42°C) in a final volume of 20 μl buffer (Tris-HCl 50 mM, KCl 75 mM, pH 8.3, Gibco BRL) containing 5 ng per μl of oligo-dT 12-18 primer (Pharmacia Biotech), 40 units of reverse transcriptase from the Moloney murine leukaemia virus (Mu-LV) (Gibco BRL, 40 units of RNAsine (Promega), DTT 10 mM (Gibco BRL) and 0.5 mM of each of the triphosphate desoxynucleotides (Promega). The cDNA samples were diluted 1/10 in distilled water and the PCR reactions were carried out using 1 μl, 4 μl or 2 μl of cDNA sample for ULIP-2 and ULIP-3 mRNAs, in a buffer (KCl 50 mM, Tris-HCl 10 mM, Triton X-100 0.1%, glycerol 0.4%, and NaCl 800 μM, pH 9), in which DTT 40 μM, $MgCl_2$ 3 mM, 0.2 mM of each dNTP, 0.4 μM of each selected primer and 2 units of Amplitaq DNA polymerase (Promega) were added in a final volume of 50 μl. The samples were then put in a thermocycler (Biomed-Hybaid), denatured at 95°C for 5 minutes and amplified during 35 cycles (one cycle = denaturation at 95°C for 65 seconds, primer hybridization 60°C for 45 seconds, extension 72°C for 4 minutes, and terminal extension 15 minutes at 72°C). The products were separated by electrophoresis on agarose-Seakem 1% gel and the bands assayed for RT-PCR products of expected size as well as the molecular

weight ladder (100 base pairs) (Promega) were visualised using ethidium bromide staining.

Composition of the oligonucleotide probes used for ULIP-3 PCR :

5'ATAGAGGAGCGGATGACG (899) 3'

GCTGTTATGGTCTTCAACTTGTCGG (1092)

GGCCTGTTATGGTCTTCAACTTGTCG (1093)

Composition of the oligonucleotide probes used for ULIP-2 PCR :

5'AGGAGGAGTGAAGACCATCG (5227) 3'

CTTATGCCACTCGCTGATGTCC (509).

B) Results

The RT-PCR experiments show that TOAD-64 (ULIP-2) and C-22 (ULIP-3)are expressed in some lung small cell tumors (see figure 8) and absent in others, such as in tumors of patients who develop paraneoplastic neurological syndromes of better prognosis.

The POP-66/ULIP-4 protein and the ULIP family proteins could be expressed in peripheric tumors (lung small cell tumor, thymome, breast and ovary cancer). Their presence could thus be correlated to a prognosis. The localisation of POP-66/ULIP-4 gene on the distal region of chromosome 10 is a confirmation in the case of brain tumors.

Thus, the differential expression of the ULIP family members in tumors such as lung small cell cancer while the corresponding ULIP gene is absent in the normal tissue, as well as the modulation of the ULIP family member expression during human HTLV1 retrovirus-mediated differentiation of a medulloblastome line, suggest that ULIP are involved in cancerous tumors.

BIBLIOGRAPHY

- Anderson et al., CRC Crit. Rev. Neurobiol., 1987, vol. 3, pp 245-99
- Antoine J.C. et al., Journal of the Neurological Sciences, 1993, vol. 117, pp 215-223
- Byk et al., Journal of Neuroscience, 1996, vol. 16(2), pp 688-701
- Dalmau et al., Neurology, 1991, vol. 41, pp 1757-64
- Duchemin et al., Dev Neurosci, 1987, vol. 9, pp 61-67
- Graus et al., Neurology, 1985, vol. 35, pp 538-543
- Graus et al., Neurology, 1990, vol. 40, pp 219-22
- Greenlee et al., Ann. Neurol., 1983, vol. 14, pp 609-13
- Hamajima et al., Gene, 1996, vol. 180, pp 157-163
- Hetzel et al., Mayo Clin. Proc., 1990, vol. 65, pp 1558-63
- Honnorat J. et al., Journal of Neurology, Neurosurgery and Psychiatry, 1996, vol. 61, pp 270-278
- Jaeckle et al., Ann. Neurol., 1985, vol. 18, pp 592-600
- Köhler and Milstein, Nature, 1975, vol. 256, pp 495-497
- Levy N., Mattei MG., 1995, Geneprobs II. A practical approach. BD Hames and SJ Higgins, Oxford University Press, pp 211-243
- Luque et al., Ann. Neurol., 1991, vol. 29, pp 241-51
- Minturn et al., J. Neurosci., 1995, vol. 15, pp 6757-6766
- Peltre G., Lapeyre J.; David B., Immunol. Lett., 1982, vol. 5, pp 127-131
- Sambrook et al., Molecular Cloning, a laboratory Manual, 1989, 9.47-9.62
- Wang L-H et al., J. Neurosci., 1996, vol. 16(9), pp 6197-6207.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Institut National de la Santé et de la Recherche Médicale

(B) STREET: 101, rue de Tolbiac

(C) CITY: Paris

(E) COUNTRY: FRANCE

(F) POSTAL CODE: 75013

(G) TELEPHONE: 0144236000

(H) TELEFAX: 0145856856

(ii) TITLE OF INVENTION: Use of ULIP in PNS and associated cancers

(iii) NUMBER OF SEQUENCES: 8

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1817 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mus musculus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CTTCCTCCCG CCCCCCGGAG AGATGTCTTA TCAGGGGAAG AAAAATATTC CACCCATCAC 60

GAGCGATCGT CTTCTGATCA AAGGTGGCAA GATTGTGAAT GATGACCAGT CCTTCTATGC 120

AGACATATAC ATGGAAGATG GGTTGATCAA GCAAATAGGA GAAAACCTGA TTGTACCAGG 180

AGGGGTGAAG ACCATCGAAG CCCACTCCAG AATGGTGATT CCCGGAGGAA TTGACGTGCA 240

TACTCGCTTC CAGATGCCTG ACCAGGGAAT GACATCCGCT GATGACTTCT TCCAGGGAAC 300

CAAGGCGGCC CTGGCCGGGG GAACCACCAT GATCATTGAC CATGTTGTTC CTGAGCCCGG 360

GACGAGCCTA TTGGCTGCCT TTGATCAGTG GAGGGAGTGG GCTGACAGCA AGTCCTGCTG 420

TGACTATTCG CTGCACGTGG ACATCACTGA GTGGCACAAG GGCATCCAGG AGGAGATGGA 480

AGCTCTGGTG AAGGACCACG GGGTAAACTC CTTCCTCGTG TACATGGCTT TCAAAGATCG 540

ATTCCAGCTG ACGGATTCCC AGATCTATGA AGTGCTGAGC GTGATCCGGG ATATCGGTGC 600

CATAGCTCAA GTCCACGCAG AGAATGGTGA CATCATTGCT GAGGCACAGC AGAGGATCCT 660

GGATCTGGGC ATCACGGGCC CCGAGGGACA CGTGTTGAGC CGGCCAGAGG AGGTCGAGGC 720

TGAAGCTGTG AACCGGTCCA TCACTATTGC CAACCAGACC AACTGCCCTC TGTATGTCAC 780

CAAAGTGATG CCCAAGAGTG CGGCTGAAGT CATCGCTCAG GCACGGAAGA AGGGAACTGT 840

GGTGTATGGT GAGCCCATCA CGGCCAGCCT GGGGACTGAT GGCTCTCATT ACTGGAGCAA 900

GAACTGGGCC AAGGCTGCGG CCTTTGTCAC CTCCCCACCC TTGAGCCCCG ACCCAACCAC 960

TCCAGACTTT CTCAACTCGT TGCTGTCCTG TGGAGACCTC CAAGTCACTG GCAGTGCCCA 1020

CTGCACCTTC AACACTGCCC AGAAGGCTGT GGGGAAGGAC AACTTCACCT TGATTCCCGA 1080

GGGCACCAAC GGCACTGAGG AGCGGATGTC TGTCATTTGG GATAAAGCTG TGGTCACTGG 1140

GAAGATGGAT GAGAATCAGT TTGTGGCTGT GACCAGCACC AACGCAGCCA AAGTCTTCAA 1200

CCTTTACCCC CGGAAAGGTC GCATCTCGGT GGGATCTGAT GCTGACTTGG TCATCTGGGA 1260

CCCTGACAGT GTGAAGACCA TCTCTGCCAA GACACACAAC AGTGCTCTTG AGTACAACAT 1320

CTTTGAAGGC ATGGAGTGTC GCGGCTCCCC ACTGGTGGTC ATCAGCCAGG GCAAGATTGT 1380

CCTGGAGGAC GGCACACTTC ATGTCACTGA AGGCTCAGGA CGCTACATTC CCCGGAAGCC 1440

CTTCCCTGAC TTTGTGTACA AACGCATCAA AGCAAGGAGC AGGCTGGCTG AGCTGAGAGG 1500

GGTCCCTCGT GGCCTGTATG ACGGACCGGT ATGCGAGGTG TCTGTGACGC CCAAGACGGT 1560

GACTCCAGCC TCATCAGCTA AGACATCCCC TGCCAAGCAG CAGGCACCAC CTGTTCGGAA 1620

CCTGCACCAG TCTGGATTCA GCTTGTCTGG TGCTCAGATT GACGACAACA TTCCCCGCCG 1680

CACCACCCAG CGCATCGTGG CACCCCCTGG TGGCCGTGCC AACATCACCA GCCTGGGCTA 1740

AAGCCCCTAG GCCTGCAGGC CACTTGGGGA TGGGGGATGG GACACCTGAG GACATTCTGA 1800

GACTTCCTTT CTTCCAT 1817

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 572 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mus musculus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Ser Tyr Gln Gly Lys Lys Asn Ile Pro Pro Ile Thr Ser Asp Arg
1               5                  10                  15

Leu Leu Ile Lys Gly Gly Lys Ile Val Asn Asp Asp Gln Ser Phe Tyr
            20                  25                  30

Ala Asp Ile Tyr Met Glu Asp Gly Leu Ile Lys Gln Ile Gly Glu Asn
        35                  40                  45

Leu Ile Val Pro Gly Gly Val Lys Thr Ile Glu Ala His Ser Arg Met
    50                  55                  60

Val Ile Pro Gly Gly Ile Asp Val His Thr Arg Phe Gln Met Pro Asp
65                  70                  75                  80

Gln Gly Met Thr Ser Ala Asp Asp Phe Phe Gln Gly Thr Lys Ala Ala
                85                  90                  95

Leu Ala Gly Gly Thr Thr Met Ile Ile Asp His Val Val Pro Glu Pro
            100                 105                 110

Gly Thr Ser Leu Leu Ala Ala Phe Asp Gln Trp Arg Glu Trp Ala Asp
        115                 120                 125

Ser Lys Ser Cys Cys Asp Tyr Ser Leu His Val Asp Ile Thr Glu Trp
    130                 135                 140

His Lys Gly Ile Gln Glu Glu Met Glu Ala Leu Val Lys Asp His Gly
145                 150                 155                 160

Val Asn Ser Phe Leu Val Tyr Met Ala Phe Lys Asp Arg Phe Gln Leu
                165                 170                 175

Thr Asp Ser Gln Ile Tyr Glu Val Leu Ser Val Ile Arg Asp Ile Gly
            180                 185                 190

Ala Ile Ala Gln Val His Ala Glu Asn Gly Asp Ile Ile Ala Glu Ala
        195                 200                 205

Gln Gln Arg Ile Leu Asp Leu Gly Ile Thr Gly Pro Glu Gly His Val
    210                 215                 220
```

```
Leu Ser Arg Pro Glu Glu Val Glu Ala Glu Ala Val Asn Arg Ser Ile
225                 230                 235                 240

Thr Ile Ala Asn Gln Thr Asn Cys Pro Leu Tyr Val Thr Lys Val Met
                245                 250                 255

Pro Lys Ser Ala Ala Glu Val Ile Ala Gln Ala Arg Lys Lys Gly Thr
            260                 265                 270

Val Val Tyr Gly Glu Pro Ile Thr Ala Ser Leu Gly Thr Asp Gly Ser
        275                 280                 285

His Tyr Trp Ser Lys Asn Trp Ala Lys Ala Ala Ala Phe Val Thr Ser
    290                 295                 300

Pro Pro Leu Ser Pro Asp Pro Thr Thr Pro Asp Phe Leu Asn Ser Leu
305                 310                 315                 320

Leu Ser Cys Gly Asp Leu Gln Val Thr Gly Ser Ala His Cys Thr Phe
                325                 330                 335

Asn Thr Ala Gln Lys Ala Val Gly Lys Asp Asn Phe Thr Leu Ile Pro
            340                 345                 350

Glu Gly Thr Asn Gly Thr Glu Glu Arg Met Ser Val Ile Trp Asp Lys
        355                 360                 365

Ala Val Val Thr Gly Lys Met Asp Glu Asn Gln Phe Val Ala Val Thr
    370                 375                 380

Ser Thr Asn Ala Ala Lys Val Phe Asn Leu Tyr Pro Arg Lys Gly Arg
385                 390                 395                 400

Ile Ser Val Gly Ser Asp Ala Asp Leu Val Ile Trp Asp Pro Asp Ser
                405                 410                 415

Val Lys Thr Ile Ser Ala Lys Thr His Asn Ser Ala Leu Glu Tyr Asn
            420                 425                 430

Ile Phe Glu Gly Met Glu Cys Arg Gly Ser Pro Leu Val Val Ile Ser
        435                 440                 445

Gln Gly Lys Ile Val Leu Glu Asp Gly Thr Leu His Val Thr Glu Gly
    450                 455                 460

Ser Gly Arg Tyr Ile Pro Arg Lys Pro Phe Pro Asp Phe Val Tyr Lys
465                 470                 475                 480

Arg Ile Lys Ala Arg Ser Arg Leu Ala Glu Leu Arg Gly Val Pro Arg
                485                 490                 495

Gly Leu Tyr Asp Gly Pro Val Cys Glu Val Ser Val Thr Pro Lys Thr
            500                 505                 510

Val Thr Pro Ala Ser Ser Ala Lys Thr Ser Pro Ala Lys Gln Gln Ala
        515                 520                 525

Pro Pro Val Arg Asn Leu His Gln Ser Gly Phe Ser Leu Ser Gly Ala
    530                 535                 540

Gln Ile Asp Asp Asn Ile Pro Arg Arg Thr Thr Gln Arg Ile Val Ala
545                 550                 555                 560

Pro Pro Gly Gly Arg Ala Asn Ile Thr Ser Leu Gly
                565                 570
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2297 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mus musculus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GCTGTCTGTC TTCAGCGCCC TCCTCTCGCC CTGCCTCTCC CTCCTCCTCC CGCCCTCCTT 60

GCCAAGCCGG GCGGTGCAGG CAGCCGGAGC AGCGGCGGCG GGCCGAGCAG CGGGGAGTGG 120

GCAGCGGTGG GAGCCGAGCT TCTGTCCTTT CTTTCATCCC TCCCTGGCCT TTGTCGCCGC 180

TCTCACGAGT AGCGCCGCCG GGAGAGACCC GGGTAGAGCG CCAGGCAGAC GTTAGTTCCA 240

GCGGCCGGGC GGAGGGCTCC AGAGGGGCCA TGTCTCATCA GGGGAAGAAG AGCATCCCGC 300

ACATCACCAG TGACCGGCTC CTCATCAGAG GTGGACGCAT CATCAATGAT GACCAGTCCT 360

TCTACGCCGA TGTCTACCTA GAAGATGGAC TCATAAAACA AATAGGAGAG AACCTGATTG 420

TTCCTGGTGG AGTGAAGACC ATCGAGGCGA ATGGCCGAAT GGTCATTCCC GGTGGCATTG 480

ATGTCAACAC TTACCTGCAG AAGCCCTCCC AGGGCATGAC CTCGGCTGAT GACTTCTTCC 540

AGGGCACTAA AGCAGCGCTG GCAGGTGGAA CCACGATGAT CATTGACCAC GTTGTTCCTG 600

AACCTGGGTC CAGCTTGTTG ACTTCCTTTG AGAAATGGCA CGAAGCAGCA GACACCAAAT 660

CCTGCTGTGA CTATTCCCTC CACGTGGACA TCACAAGCTG GTATGATGGT GTTCGGGAAG 720

AGCTGGAGGT GCTGGTGCAG GACAAAGGTG TCAACTCCTT CCAAGTCTAC ATGGCGTATA 780

AGGACCTGTA CCAGATGTCT GACAGCCAGC TGTATGAAGC CTTCACCTTC CTTAAGGGTT 840

TGGGAGCTGT GATCTTAGTC CATGCAGAAA ATGGAGATTT GATAGCTCAG GAACAAAAAC 900

GGATCCTGGA GATGGGCATC ACGGGTCCCG AGGGTCATGC TCTGAGCAGA CCCGAGGAGC 960

TGGAGGCCGA GGCTGTGTTC CGGGCTATTG CCATTGCAGG CCGGATCAAT TGCCCTGTGT 1020

ACATCACCAA GGTCATGAGC AAGAGTGCAG CGGACATCAT CGCACTGGCC AGGAAGAAAG 1080

GCCCTCTTGT CTTCGGTGAG CCCATAGCCG CCAGCCTGGG AACCGATGGC ACCCACTACT 1140

GGAGCAAGAA CTGGGCCAAG GCAGCTGCAT TTGTGACTTC CCCTCCCCTG AGCCCAGACC 1200

CCACCACTCC TGACTACTTG ACCTCCTTGC TGGCCTGTGG AGACTTGCAG GTCACAGGTA 1260

GTGGCCACTG TCCCTACAGT ATTGCTCAGA AGGCTGTGGG CAAGGACAAC TTCACTCTGA 1320

TCCCTGAGGG TGTCAATGGT ATAGAAGAGC GGATGACCGT TGTCTGGGAC AAGGCAGTGG 1380

CTACTGGCAA GATGGATGAG AACCAGTTTG TAGCCGTCAC CAGCACCAAC GCAGCCAAGA 1440

TCTTCAACCT GTACCCGAGG AAAGGTCGGA TCGCTGTGGG CTCCGATGCT GACGTAGTCA 1500

TCTGGGACCC AGATAAGATG AAGACCATAA CAGCCAAAAG CCATAAATCA ACTGTGGAGT 1560

ACAACATCTT TGAGGGCATG GAGTGCCACG GCTCCCCCCT GGTGGTCATC AGTCAGGGCA 1620

AGATTGTCTT TGAGGATGGA AACATCAGTG TCAGCAAGGG CATGGGCCGC TTCATCCCTC 1680

GGAAGCCATT CCCAGAGCAT CTCTACCAGC GTGTCAGGAT CAGAAGCAAG GTTTTCGGGT 1740

TGCATAGTGT TTCCAGGGGC ATGTACGATG GGCCTGTGTA CGAGGTGCCA GCTACACCCA 1800

AACATGCTGC TCCTGCTCCT TCTGCCGAAT CCTCGCCTTC TAAACACCAA CCCCCACCCA 1860

TCCGGAACCT CCACCAGTCC AACTTCAGCT TATCAGGTGC CCAGATAGAT GACAACAATC 1920

CAAGGCGTAC AGGCCACCGC ATTGTGGCGC CCCCTGGTGG CCGCTCCAAC ATCACCAGCC 1980

TCGGTTGACC TCAGATGAGC CAGATATGCA AGAGTGAAGG ATTATGGGAA AACGTCCATT 2040

CCTTTTCCGT GTTTTTGAAG CCCACAGTTT TAGTTGGTAC TGACGGAGGG GAGATTGAGC 2100

GATGCTCTTT CCTTCTCTGT TTAGGAAGAA GTGGTACTAG TGTGGTGTGT TTGCCTGGAA 2160

GTCCCTCGCC CACAGTGTGT GTTCACACCG ACTCCACCTC AGAGCATGGT GCCGTCCGTT 2220

TTCCCTTCCT AGTGACCCCA GGTTTAGCAT CGTCCTATAC TGTTCCCTCC ACTCCTCCAT 2280

GACCCTCTGA GTGATGG 2297

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 572 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mus musculus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Ser His Gln Gly Lys Lys Ser Ile Pro His Ile Thr Ser Asp Arg
1               5                  10                  15

Leu Leu Ile Arg Gly Gly Arg Ile Ile Asn Asp Asp Gln Ser Phe Tyr
            20                  25                  30

Ala Asp Val Tyr Leu Glu Asp Gly Leu Ile Lys Gln Ile Gly Glu Asn
        35                  40                  45

Leu Ile Val Pro Gly Gly Val Lys Thr Ile Glu Ala Asn Gly Arg Met
    50                  55                  60

Val Ile Pro Gly Gly Ile Asp Val Asn Thr Tyr Leu Gln Lys Pro Ser
65                  70                  75                  80

Gln Gly Met Thr Ser Ala Asp Asp Phe Phe Gln Gly Thr Lys Ala Ala
                85                  90                  95

Leu Ala Gly Gly Thr Thr Met Ile Ile Asp His Val Val Pro Glu Pro
            100                 105                 110

Gly Ser Ser Leu Leu Thr Ser Phe Glu Lys Trp His Glu Ala Ala Asp
        115                 120                 125

Thr Lys Ser Cys Cys Asp Tyr Ser Leu His Val Asp Ile Thr Ser Trp
    130                 135                 140

Tyr Asp Gly Val Arg Glu Glu Leu Glu Val Leu Val Gln Asp Lys Gly
145                 150                 155                 160

Val Asn Ser Phe Gln Val Tyr Met Ala Tyr Lys Asp Leu Tyr Gln Met
                165                 170                 175

Ser Asp Ser Gln Leu Tyr Glu Ala Phe Thr Phe Leu Lys Gly Leu Gly
            180                 185                 190

Ala Val Ile Leu Val His Ala Glu Asn Gly Asp Leu Ile Ala Gln Glu
        195                 200                 205

Gln Lys Arg Ile Leu Glu Met Gly Ile Thr Gly Pro Glu Gly His Ala
    210                 215                 220
```

```
Leu Ser Arg Pro Glu Glu Leu Glu Ala Glu Ala Val Phe Arg Ala Ile
225                 230                 235                 240

Ala Ile Ala Gly Arg Ile Asn Cys Pro Val Tyr Ile Thr Lys Val Met
                245                 250                 255

Ser Lys Ser Ala Ala Asp Ile Ile Ala Leu Ala Arg Lys Lys Gly Pro
            260                 265                 270

Leu Val Phe Gly Glu Pro Ile Ala Ala Ser Leu Gly Thr Asp Gly Thr
        275                 280                 285

His Tyr Trp Ser Lys Asn Trp Ala Lys Ala Ala Ala Phe Val Thr Ser
    290                 295                 300

Pro Pro Leu Ser Pro Asp Pro Thr Thr Pro Asp Tyr Leu Thr Ser Leu
305                 310                 315                 320

Leu Ala Cys Gly Asp Leu Gln Val Thr Gly Ser Gly His Cys Pro Tyr
                325                 330                 335

Ser Ile Ala Gln Lys Ala Val Gly Lys Asp Asn Phe Thr Leu Ile Pro
            340                 345                 350

Glu Gly Val Asn Gly Ile Glu Glu Arg Met Thr Val Val Trp Asp Lys
        355                 360                 365

Ala Val Ala Thr Gly Lys Met Asp Glu Asn Gln Phe Val Ala Val Thr
    370                 375                 380

Ser Thr Asn Ala Ala Lys Ile Phe Asn Leu Tyr Pro Arg Lys Gly Arg
385                 390                 395                 400

Ile Ala Val Gly Ser Asp Ala Asp Val Val Ile Trp Asp Pro Asp Lys
                405                 410                 415

Met Lys Thr Ile Thr Ala Lys Ser His Lys Ser Thr Val Glu Tyr Asn
            420                 425                 430

Ile Phe Glu Gly Met Glu Cys His Gly Ser Pro Leu Val Val Ile Ser
        435                 440                 445

Gln Gly Lys Ile Val Phe Glu Asp Gly Asn Ile Ser Val Ser Lys Gly
    450                 455                 460

Met Gly Arg Phe Ile Pro Arg Lys Pro Phe Pro Glu His Leu Tyr Gln
465                 470                 475                 480

Arg Val Arg Ile Arg Ser Lys Val Phe Gly Leu His Ser Val Ser Arg
                485                 490                 495

Gly Met Tyr Asp Gly Pro Val Tyr Glu Val Pro Ala Thr Pro Lys His
            500                 505                 510

Ala Ala Pro Ala Pro Ser Ala Glu Ser Ser Pro Ser Lys His Gln Pro
        515                 520                 525

Pro Pro Ile Arg Asn Leu His Gln Ser Asn Phe Ser Leu Ser Gly Ala
    530                 535                 540

Gln Ile Asp Asp Asn Asn Pro Arg Arg Thr Gly His Arg Ile Val Ala
545                 550                 555                 560

Pro Pro Gly Gly Arg Ser Asn Ile Thr Ser Leu Gly
                565                 570
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1920 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mus musculus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GCTGACTAAT ATGCTTAAAT TCAGCGGGTC GCCACGTCTG GTCGGTACGT CCACGCCCGC 60
GCAGCCCCTA CCGAGGACAC TCAGCCCGCC CGTGTATCAG GATGTCCTTC CAAGGCAAGA 120
AGAGCATTCC CCGGATAACG AGCGACCGCC TTCTCATCAA AGGTGGGAAG ATTGTGAACG 180
ATGACCAGTC CTTTCATGCT GATCTGTATG TGGAAGACGG TCTGATTAAA CAAATTGGAG 240
AAAATCTCAT CGTCCCTGGG GGCATCAAAA CCATCGATGC TCATGGCCTG ATGGTGCTGC 300
CTGGGGGAGT TGACGTTCAC ACCCGGCTGC AGATGCCTGT GATGGGCATG ACCCCAGCTG 360
ATGATTTCTG TCAGGGCACC AAGGCGGCTC TAGCAGGCGG GACCACCATG ATATTGGACC 420
ATGTGTTTCC TGACGCTGGT GTGAGCCTGC TGGCAGCCTA TGAGCAGTGG CGGGACGGAG 480
CAGACAGCGC GGCCTGCTGT GACTACTCCT TACATGTGGA CATTCCTCGC TGGCACGAGA 540
GCACCAAAGA AGAGCTGGAG GCCCTAGTCA GGGACAAAGG TGTGAACTCC TTCCTGGTCT 600
TCATGGCATA CAAGGACAGG TGCCAGTGTA CTGACGGTCA GATATATGAA ATCTTCAGCC 660
TCATCCGGGA CCTGGGAGCT GTGGCCCAGG TGCACGCAGA AAATGGGGAC ATCGTGGAGG 720
AGGAACAGAA GCGCCTGCTG GAGCAAGGCA TCACTGGTCC TGAGGGCCAT GTGCTCAGCC 780
ACCCAGAAGA GGTAGAGGCC GAGGCTGTGT ACAGAGCAGT CACCATTGCC AAGCAGGCCA 840
ACTGCCCACT ATACGTCACC AAGGTGATGA GCAAGGGTGC AGCTGACATG GTTGCCCAAG 900
CCAAGCGCAG GGGGGTGGTC GTCTTTGGGG AACCTATCAC TGCCAGCCTG GGCACTGATG 960
GCTCACACTA CTGGAGCAAG AACTGGGCCA AGGCTGCAGC CTTTGTCACT TCACCCCCTA 1020
TCAACCCGGA CCCTACTACT GCAGACCACC TCACCTCTCT GCTGTCCAGT GGGGACCTCC 1080
AGGTGACAGG CAGTGCCCAC TGCACCTTCA CTACTGCCCA GAAGGCTGTT GGCAAAGACA 1140

ACTTCACACT GATCCCCGAG GTAGTCAACG GTATAGAAGA GCGCATGTCT GTGGTCTGGG 1200

AGAAATGTGT GGCTTCAGGG AAAATGGACG AGAATGAGTT CGTTGCCGTG ACCAGCACAA 1260

ATGCTGCCAA AATCTTCAAT TTTTACCCCA GGAAGGGGCG TGTGGCCGTG GGCTCTGATG 1320

CTGACCTGGT CATCTGGAAC CCCAGGGCCA CGAAAGTCAT CTCTGCCAAG AGCCATAACC 1380

TGAATGTAGA GTACAACATC TTTGAAGGAG TGGAGTGCCG AGGAGTGCCC ACGGTGGTCA 1440

TAAGTCAGGG CAGAGTGGTG CTGGAGGACG GAAACCTGCT TGTCACTCCA GGGGCTGGCC 1500

GCTTCATTCC CCGGAAGACG TTCCCGGACT TTGTCTATAA GAGGATAAAG GCTCGCAACA 1560

GGCTAGCAGA GATCCACGGT GTGCCTCGAG GCCTGTACGA CGGGCCTGTG CATGAAGTGA 1620

TGTTACCTGC CAAGCCAGGA AGTGGCACAC AGGCCCGTGC ATCCTGTTCA GGCAAGATCT 1680

CAGTGCCACC CGTGCGCAAC CTGCACCAGT CGGGGTTCAG CCTATCTGGC TCTCAGGCTG 1740

ACGATCACAT TGCCAGACGT ACGGCTCAGA AGATCATGGC ACCCCCCGGA GGACGCTCCA 1800

ACATCACGTC TCTTTCCTAG ACTTGGGGTC TTGGCAAGCT GGTGCTGTCC CCACTGGCAG 1860

GGTGTGGGGA CGACTCACGT CAGTTAGCTC CTTCCTTTGT AGATTGTTAT TGTGAAAGGC 1920

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 572 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mus musculus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Ser Phe Gln Gly Lys Lys Ser Ile Pro Arg Ile Thr Ser Asp Arg
1               5                  10                  15

Leu Leu Ile Lys Gly Gly Lys Ile Val Asn Asp Asp Gln Ser Phe His
            20                  25                  30

Ala Asp Leu Tyr Val Glu Asp Gly Leu Ile Lys Gln Ile Gly Glu Asn
        35                  40                  45

Leu Ile Val Pro Gly Gly Ile Lys Thr Ile Asp Ala His Gly Leu Met
    50                  55                  60

Val Leu Pro Gly Gly Val Asp Val His Thr Arg Leu Gln Met Pro Val
65                  70                  75                  80

Met Gly Met Thr Pro Ala Asp Asp Phe Cys Gln Gly Thr Lys Ala Ala
                85                  90                  95

Leu Ala Gly Gly Thr Thr Met Ile Leu Asp His Val Phe Pro Asp Ala
            100                 105                 110

Gly Val Ser Leu Leu Ala Ala Tyr Glu Gln Trp Arg Asp Gly Ala Asp
        115                 120                 125

Ser Ala Ala Cys Cys Asp Tyr Ser Leu His Val Asp Ile Pro Arg Trp
    130                 135                 140

His Glu Ser Thr Lys Glu Glu Leu Glu Ala Leu Val Arg Asp Lys Gly
145                 150                 155                 160

Val Asn Ser Phe Leu Val Phe Met Ala Tyr Lys Asp Arg Cys Gln Cys
                165                 170                 175

Thr Asp Gly Gln Ile Tyr Glu Ile Phe Ser Leu Ile Arg Asp Leu Gly
            180                 185                 190

Ala Val Ala Gln Val His Ala Glu Asn Gly Asp Ile Val Glu Glu Glu
        195                 200                 205

Gln Lys Arg Leu Leu Glu Gln Gly Ile Thr Gly Pro Glu Gly His Val
    210                 215                 220

Leu Ser His Pro Glu Glu Val Glu Ala Glu Ala Val Tyr Arg Ala Val
225                 230                 235                 240

Thr Ile Ala Lys Gln Ala Asn Cys Pro Leu Tyr Val Thr Lys Val Met
                245                 250                 255
```

```
Ser Lys Gly Ala Ala Asp Met Val Ala Gln Ala Lys Arg Arg Gly Val
            260                 265                 270

Val Val Phe Gly Glu Pro Ile Thr Ala Ser Leu Gly Thr Asp Gly Ser
        275                 280                 285

His Tyr Trp Ser Lys Asn Trp Ala Lys Ala Ala Ala Phe Val Thr Ser
    290                 295                 300

Pro Pro Ile Asn Pro Asp Pro Thr Thr Ala Asp His Leu Thr Ser Leu
305                 310                 315                 320

Leu Ser Ser Gly Asp Leu Gln Val Thr Gly Ser Ala His Cys Thr Phe
                325                 330                 335

Thr Thr Ala Gln Lys Ala Val Gly Lys Asp Asn Phe Thr Leu Ile Pro
            340                 345                 350

Glu Val Val Asn Gly Ile Glu Glu Arg Met Ser Val Val Trp Glu Lys
        355                 360                 365

Cys Val Ala Ser Gly Lys Met Asp Glu Asn Glu Phe Val Ala Val Thr
    370                 375                 380

Ser Thr Asn Ala Ala Lys Ile Phe Asn Phe Tyr Pro Arg Lys Gly Arg
385                 390                 395                 400

Val Ala Val Gly Ser Asp Ala Asp Leu Val Ile Trp Asn Pro Arg Ala
                405                 410                 415

Thr Lys Val Ile Ser Ala Lys Ser His Asn Leu Asn Val Glu Tyr Asn
            420                 425                 430

Ile Phe Glu Gly Val Glu Cys Arg Gly Val Pro Thr Val Val Ile Ser
        435                 440                 445

Gln Gly Arg Val Val Leu Glu Asp Gly Asn Leu Leu Val Thr Pro Gly
    450                 455                 460

Ala Gly Arg Phe Ile Pro Arg Lys Thr Phe Pro Asp Phe Val Tyr Lys
465                 470                 475                 480

Arg Ile Lys Ala Arg Asn Arg Leu Ala Glu Ile His Gly Val Pro Arg
                485                 490                 495

Gly Leu Tyr Asp Gly Pro Val His Glu Val Met Leu Pro Ala Lys Pro
            500                 505                 510

Gly Ser Gly Thr Gln Ala Arg Ala Ser Cys Ser Gly Lys Ile Ser Val
        515                 520                 525

Pro Pro Val Arg Asn Leu His Gln Ser Gly Phe Ser Leu Ser Gly Ser
    530                 535                 540

Gln Ala Asp Asp His Ile Ala Arg Arg Thr Ala Gln Lys Ile Met Ala
545                 550                 555                 560

Pro Pro Gly Gly Arg Ser Asn Ile Thr Ser Leu Ser
                565                 570
```

INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCCGCCCCTA CCAGAGACCC CCAGGAGCAG GATGTCCTTC CAGGGCAAGA AAAGCATCCC 60
CCGGATCACG AGTGACCGCC TTCTGATCAG AGGTGGGAGG ATCGTGAATG ACGACCAGTC 120
CTTTTACGCT GATGTGCACG TGGAAGATGG CTTGATAAAA CAAATCGGAG AAAACCTCAT 180
CGTCCCTGGG GGCATCAAGA CCATTGACGC CCACGGCCTG ATGGTCCTTC CTGGTGGCGT 240
TGACGTCCAC ACAAGGCTGC AGATGCCTGT CCTGGGCATG ACACCGGCTG ACGACTTCTG 300
TCAGGGCACC AAGGCAGCGC TAGCAGGAGG AACCACCATG ATCTTGGACC ACGTCTTCCC 360
CGACACGGGT GTGAGCCTGC TGGCGGCCTA CGAGCAGTGG CGGGAGCGGG CGGACAGCGC 420
GGCCTGCTGC GACTACTCCC TGCACGTGGA CATCACCCGA TGGCATGAGA GCATCAAGGA 480
GGAGCTGGAG GCCCTGGTCA AGGAGAAGGG TGTGAACTCC TTCCTGGTCT TCATGGCATA 540
CAAGGACCGG TGCCAGTGCA GCGACAGCCA GATGTACGAG ATCTTCAGCA TCATCCGGGA 600
CCTGGGGGCC TTGGCCCAGG TGCACGCTGA GAACGGGGAC ATCGTGGAGG AGGAGCAGAA 660
GCGGTTGCTG GAGCTCGGCA TCACTGGCCC CGAGGGCCAC GTGCTCAGCC ACCCCGAGGA 720
GGTGGAGGCT GAGGCGGTGT ACCGAGCTGT CACCATCGCC AAGCAGGCAA ACTGCCCGCT 780
GTACGTCACC AAGGTGATGA GCAAGGGGGC GGCCGACGCC ATCGCTCAGG CCAAGCGCAG 840
AGGGGTGGTC GTGTTTGGGG AGCCCATCAC CGCCAGCCTG GGCACCGACG GTTCACACTA 900
CTGGAGCAAG AACTGGGCCA AGGCTGCAGC CTTCGTCACA TCACCCCCTG TCAACCCAGA 960
CCCCACCACG GCAGACCACC TCACCTGCTT GCTGTCCAGC GGGGACCTCC AGGTGACAGG 1020
CAGCGCCCAC TGCACCTTCA CCACTGCCCA GAAGGCTGTG GGCAAGGACA ACTTCGCGCT 1080
GATCCCCGAG GGCACCAACG GCATTGAGGA GCGCATGTCG ATGGTCTGGG AGAAATGTGT 1140

GGCCTCTGGG AAGATGGACG AGAATGAGTT CGTCGCGGTG ACCAGTACAA ATGCTGCCAA 1200

AATCTTCAAT TTTTACCCAA GGAAGGGGCG AGTGGCTGTG GGCTCTGACG CTGACCTGGT 1260

CATATGGAAC CCCAAGGCCA CCAAGATCAT CTCTGCCAAG ACCCACAATC TGAACGTGGA 1320

GTACAACATC TTCGAGGGAG TGGAGTGCCG GGGAGCGCCT GCCGTGGTCA TAAGTCAGGG 1380

CCGAGTGGCG CTGGAGGACG GGAAGATGTT TGTCACCCCG GGGGCGGGCC GCTTCGTCCC 1440

TCGGAAAACA TTCCCGGACT TTGTCTACAA GAGGATCAAA GCTCGCAACA GGCTGGCGGA 1500

GATCCACGGT GTGCCCCGTG GGCTGTATGA CGGGCCCGTC CACGAGGTGA TGGTGCCTGC 1560

CAAGCCAGGG AGTGGCGCTC CGGCCCGCGC GTCCTGCCCA GGCAAGATCT CCGTGCCTCC 1620

TGTGCGCAAC CTACATCAGT CGGGGTTCAG CCTATCTGGG TCTCAGGCTG ATGACCACAT 1680

CGCCCGACGC 1690

INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 572 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Met Ser Phe Gln Gly Lys Lys Ser Ile Pro Arg Ile Thr Ser Asp Arg
1               5                  10                  15

Leu Leu Ile Arg Gly Gly Arg Ile Val Asn Asp Asp Gln Ser Phe Tyr
            20                  25                  30

Ala Asp Val His Val Glu Asp Gly Leu Ile Lys Gln Ile Gly Glu Asn
        35                  40                  45

Leu Ile Val Pro Gly Gly Ile His Thr Ile Asp Ala His Gly Leu Met
    50                  55                  60

Val Leu Pro Gly Gly Val Asp Val His Thr Arg Leu Gln Met Pro Val
65                  70                  75                  80

Leu Gly Met Thr Pro Ala Asp Asp Phe Cys Gln Gly Thr Lys Ala Ala
                85                  90                  95

Leu Ala Gly Gly Thr Thr Met Ile Leu Asp His Val Phe Pro Asp Thr
            100                 105                 110

Gly Val Ser Leu Leu Ala Ala Tyr Glu Gln Trp Arg Glu Arg Ala Asp
        115                 120                 125

Ser Ala Ala Cys Cys Asp Tyr Ser Leu His Val Asp Ile Thr Arg Trp
    130                 135                 140

His Glu Ser Ile Lys Glu Glu Leu Glu Ala Leu Val Lys Glu Lys Gly
145                 150                 155                 160

Val Asn Ser Phe Leu Val Phe Met Ala Tyr Lys Asp Arg Cys Gln Cys
                165                 170                 175

Ser Asp Ser Gln Met Tyr Glu Ile Phe Ser Ile Ile Arg Asp Leu Gly
            180                 185                 190

Ala Leu Ala Gln Val His Ala Glu Asn Gly Asp Ile Val Glu Glu Glu
        195                 200                 205

Gln Lys Arg Leu Leu Glu Leu Gly Ile Thr Gly Pro Glu Gly His Val
    210                 215                 220

Leu Ser His Pro Glu Glu Val Glu Ala Glu Ala Val Tyr Arg Ala Val
225                 230                 235                 240

Thr Ile Ala Lys Gln Ala Asn Cys Pro Leu Tyr Val Thr Lys Val Met
                245                 250                 255
```

Ser Lys Gly Ala Ala Asp Ala Ile Ala Gln Ala Lys Arg Arg Gly Val
260 265 270

Val Val Phe Gly Glu Pro Ile Thr Ala Ser Leu Gly Thr Asp Gly Ser
275 280 285

His Tyr Trp Ser Lys Asn Trp Ala Lys Ala Ala Ala Phe Val Thr Ser
290 295 300

Pro Pro Val Asn Pro Asp Pro Thr Thr Ala Asp His Leu Thr Cys Leu
305 310 315 320

Leu Ser Ser Gly Asp Leu Gln Val Thr Gly Ser Ala His Cys Thr Phe
325 330 335

Thr Thr Ala Gln Lys Ala Val Gly Lys Asp Asn Phe Ala Leu Ile Pro
340 345 350

Glu Gly Thr Asn Gly Ile Glu Glu Arg Met Ser Met Val Trp Glu Lys
355 360 365

Cys Val Ala Ser Gly Lys Met Asp Glu Asn Glu Phe Val Ala Val Thr
370 375 380

Ser Thr Asn Ala Ala Lys Ile Phe Asn Phe Tyr Pro Arg Lys Gly Arg
385 390 395 400

Val Ala Val Gly Ser Asp Ala Asp Leu Val Ile Trp Asn Pro Lys Ala
405 410 415

Thr Lys Ile Ile Ser Ala Lys Thr His Asn Leu Asn Val Glu Tyr Asn
420 425 430

Ile Phe Glu Gly Val Glu Cys Arg Gly Ala Pro Ala Val Val Ile Ser
435 440 445

Gln Gly Arg Val Ala Leu Glu Asp Gly Lys Met Phe Val Thr Pro Gly
450 455 460

Ala Gly Arg Phe Val Pro Arg Lys Thr Phe Pro Asp Phe Val Tyr Lys
465 470 475 480

Arg Ile Lys Ala Arg Asn Arg Leu Ala Glu Ile His Gly Val Pro Arg
485 490 495

Gly Leu Tyr Asp Gly Pro Val His Glu Val Met Val Pro Ala Lys Pro
500 505 510

Gly Ser Gly Ala Pro Ala Arg Ala Ser Cys Pro Gly Lys Ile Ser Val
515 520 525

Pro Pro Val Arg Asn Leu His Gln Ser Gly Phe Ser Leu Ser Gly Ser
530 535 540

Gln Ala Asp Asp His Ile Ala Arg Arg Thr Ala Gln Lys Ile Met Ala
545 550 555 560

Pro Pro Gly Gly Arg Ser Asn Ile Thr Ser Leu Ser
565 570

CLAIMS

1. Purified polypeptide, derivative or polypeptide fragment of the said biologically active purified polypeptide, comprising an amino acid sequence selected from SEQ ID No. 2, No. 4, No. 6 and No. 8.

2. Purified polypeptide, derivative or biologically active polypeptide fragment of the said purified polypeptide, according to Claim 1, comprising the amino acid sequence SEQ ID No. 8, the said polypeptide being designated by "POP-66".

3. Isolated nucleotide sequence, comprising:

- a sequence selected from SEQ ID No. 1, No. 3, No. 5 and No. 7 coding for a polypeptide of amino acid sequence SEQ ID No. 2, No. 4, No. 6 and No. 8 respectively;

- a sequence derived from a sequence selected from SEQ ID No. 1, No. 3, No. 5 and No. 7 on account of the degeneracy of the genetic code.

4. Nucleotide sequence according to Claim 3, comprising the nucleotide sequence SEQ ID No. 7 coding for a polypeptide according to Claim 2.

5. Cloning and/or expression vector containing a nucleic acid sequence according to one of Claims 3 and 4.

6. Host cell transfected by a vector according to Claim 5.

7. Mono- or polyclonal antibodies obtained from a purified polypeptide according to one of Claims 1 and 2, a derivative or a biologically active polypeptide fragment of the said purified polypeptide, as well as the fragments, the chimeric antibodies or the immunoconjugates of the said mono- or polyclonal antibodies.

8. Composition useful for the diagnosis of paraneoplastic neurological syndromes and/or for the early diagnosis of the formation of tumours, characterized in that it comprises a purified POP-66

polypeptide, derivative or biologically active polypeptide fragment of POP-66 according to Claim 2.

9. Use of a purified POP-66 polypeptide, derivative or biologically active polypeptide fragment of POP-66 according to Claim 2 or of a nucleotide sequence according to Claim 4 for detecting the presence of anti-CV2 antibodies in a biological sample.

10. Use of mono- or polyclonal antibodies or their fragments, chimeric or immunoconjugated antibodies according to Claim 7 for the purification or the detection of a corresponding ULIP protein in a biological sample.

11. Use of antibodies directed against a protein of the ULIP family for the demonstration of a ULIP protein in neoplasms and paraneoplastic neurological syndromes, for diagnostic purposes.

12. Use according to Claim 11, the antibodies being monoclonal antibodies obtained from polyclonal anti-CV2 serum of patients.

13. Method for the diagnosis of paraneoplastic neurological syndromes and/or for the early diagnosis of the formation of cancerous tumours, characterized in that auto-antibodies directed against a POP-66 protein are demonstrated in a blood sample taken from an individual by

- the contacting of a blood sample taken from an individual with a purified polypeptide (POP-66), derivative or biologically active polypeptide fragment of POP-66 according to Claim 2, optionally attached to a support under conditions allowing the formation of specific immunological complexes between the said polypeptide and the auto-antibodies optionally present in the blood sample, and

- the detection of the specific immunological complexes optionally formed.

14. Kit for the diagnosis of paraneoplastic neurological syndromes and for the early diagnosis of

the formation of tumours from a biological sample, comprising:

- at least one purified POP-66 polypeptide, derivative or biologically active polypeptide fragment of POP-66, according to Claim 2, optionally attached to a support,

- means of visualization of the formation of specific antigen/antibody complexes between an anti-POP-66 auto-antibody and the said purified POP-66 polypeptide, derivative or polypeptide fragment and/or means of quantification of these complexes.

15. Pharmaceutical composition, comprising at least one purified protein of the ULIP family, polypeptide fragment or biologically active derivative of this, a nucleotide sequence or nucleotide sequence fragment coding for the said protein, an antisense sequence capable of hybridizing specifically with a nucleotide sequence coding for the said protein, or an antibody directed against the said protein, combined with a pharmaceutically acceptable vehicle.

16. Pharmaceutical composition according to Claim 15, comprising at least one purified POP-66 polypeptide according to Claim 2, polypeptide fragment or biologically active derivative of this, a nucleotide sequence or nucleotide sequence fragment coding for the said polypeptide, an antisense sequence capable of hybridizing specifically with a nucleotide sequence coding for the said polypeptide, or an antibody directed against the said polypeptide, combined with a pharmaceutically acceptable vehicle.

17. Use of a purified protein of the ULIP family, polypeptide fragment or biologically active derivative of this, a nucleotide sequence or nucleotide sequence fragment coding for the said protein, an antisense sequence capable of hybridizing specifically with a nucleotide sequence coding for the said protein, or an antibody directed against the said protein, for the

production of a medicament intended for treating neurodegenerative illnesses and neoplasms.

FIG. 1A

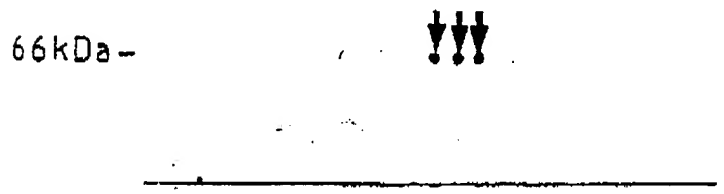


FIG. 1B

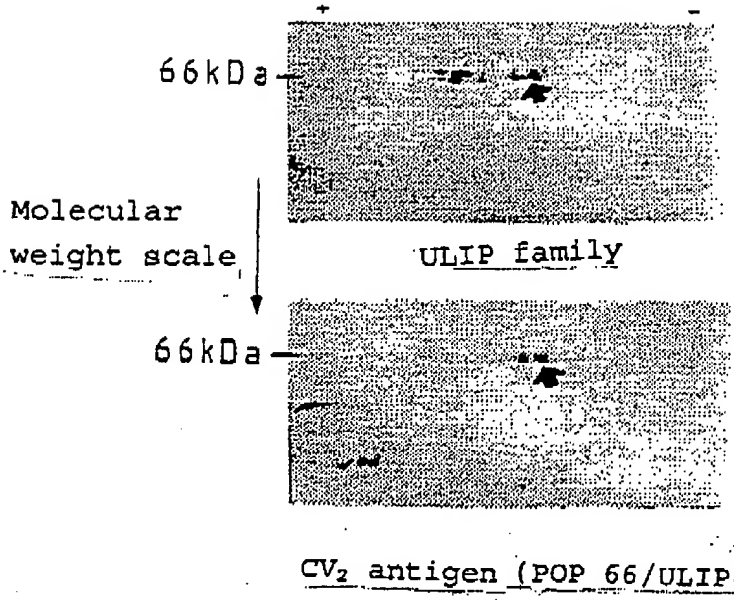


FIG. 2A

FIG. 2B

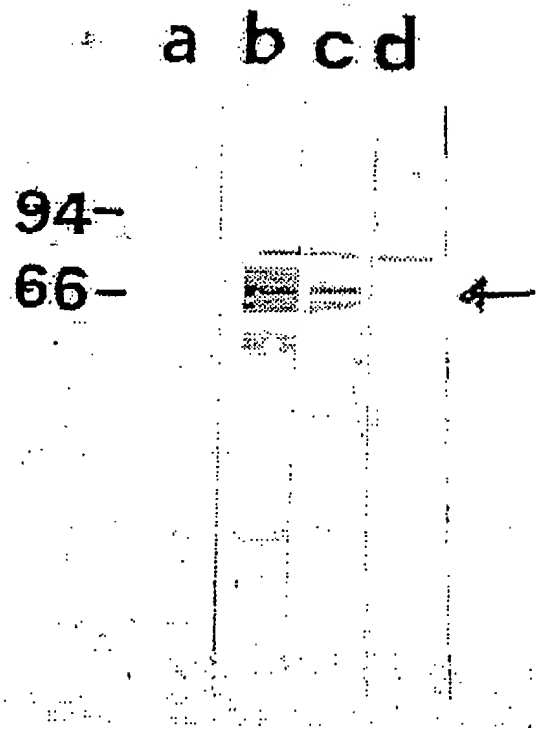


FIG. 3

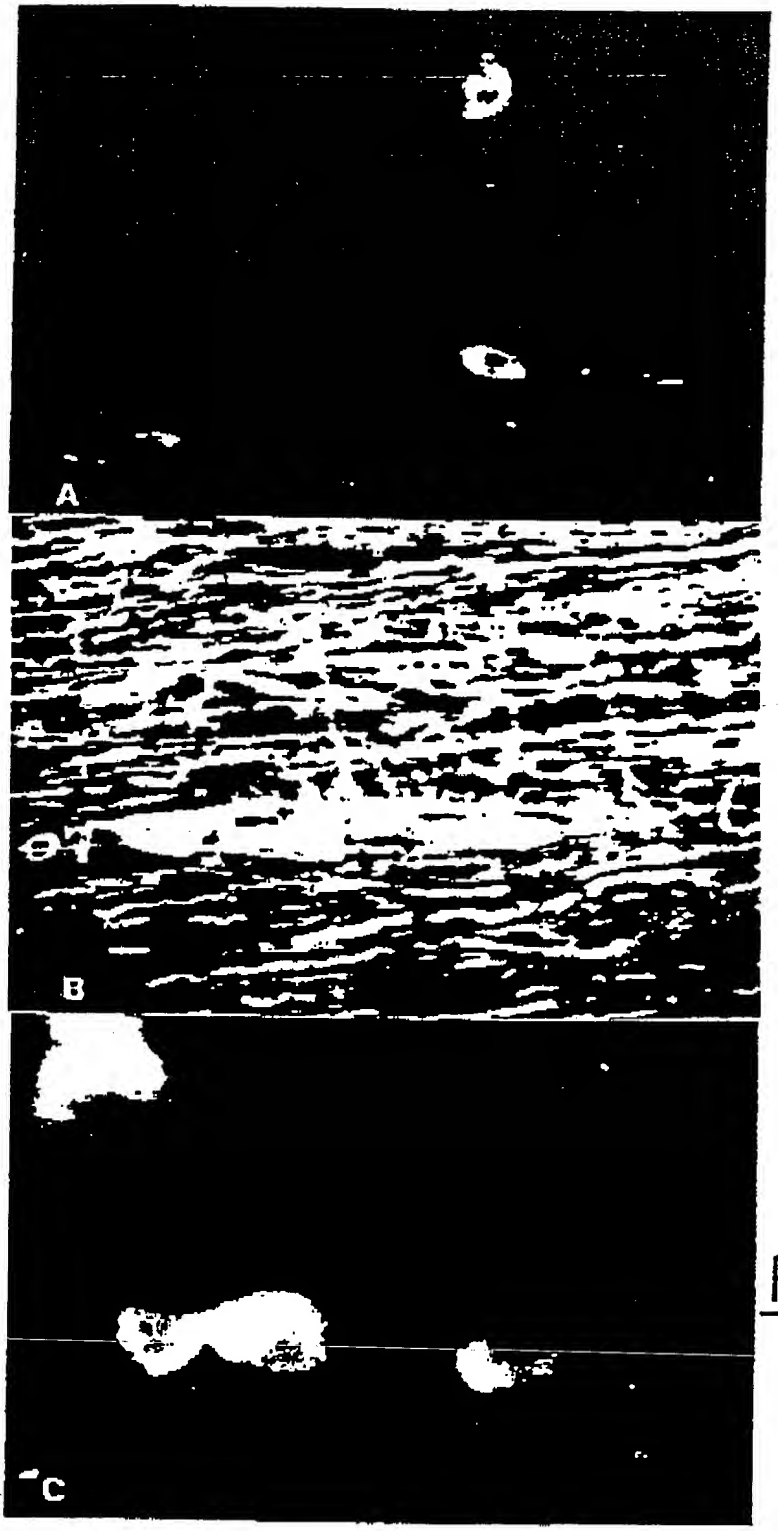


FIG.4A

FIG.4B

FIG.4C

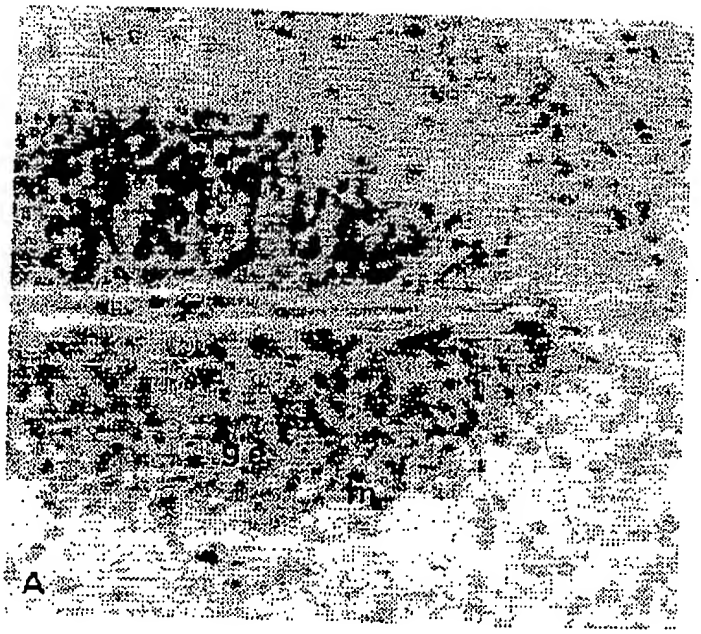


FIG. 5A

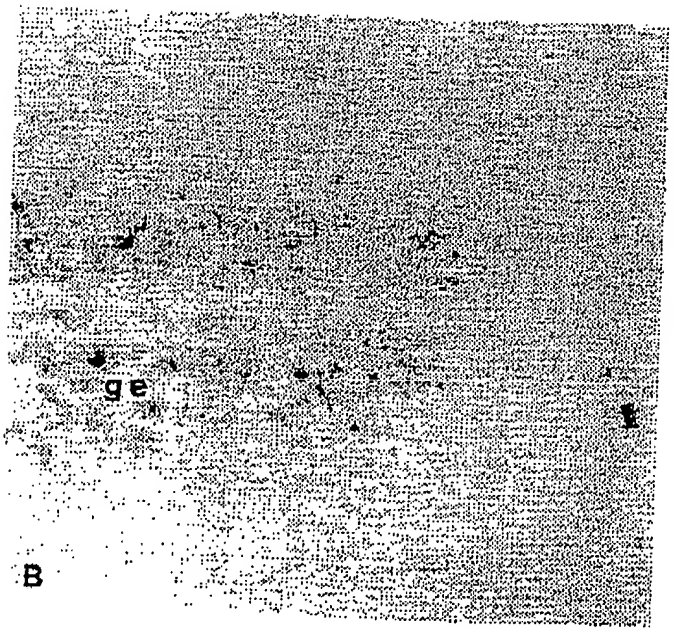


FIG. 5B

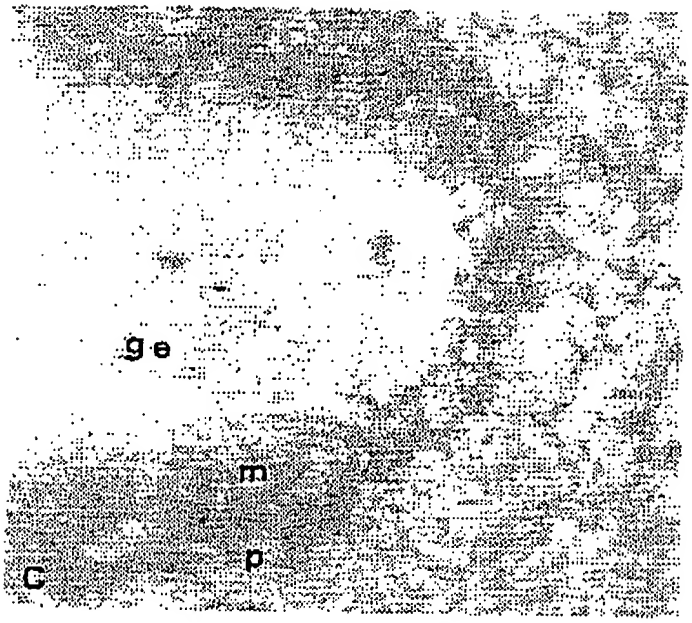


FIG. 5C

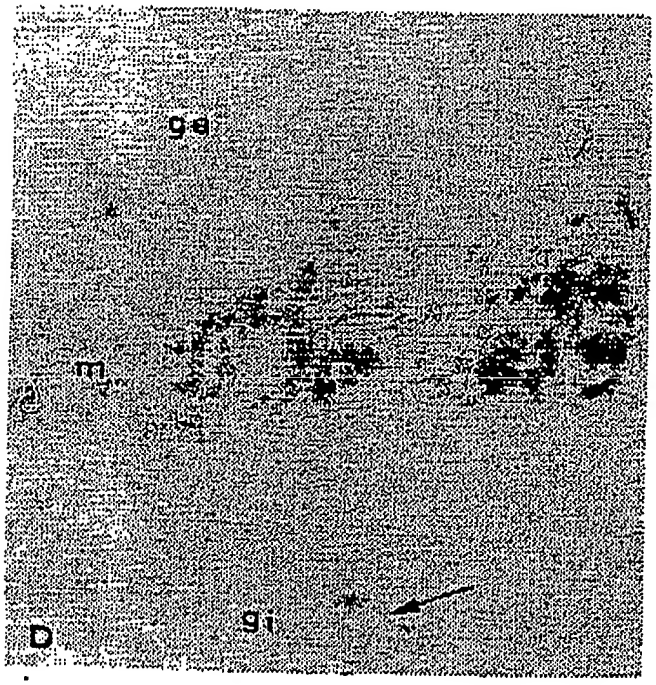


FIG. 5D

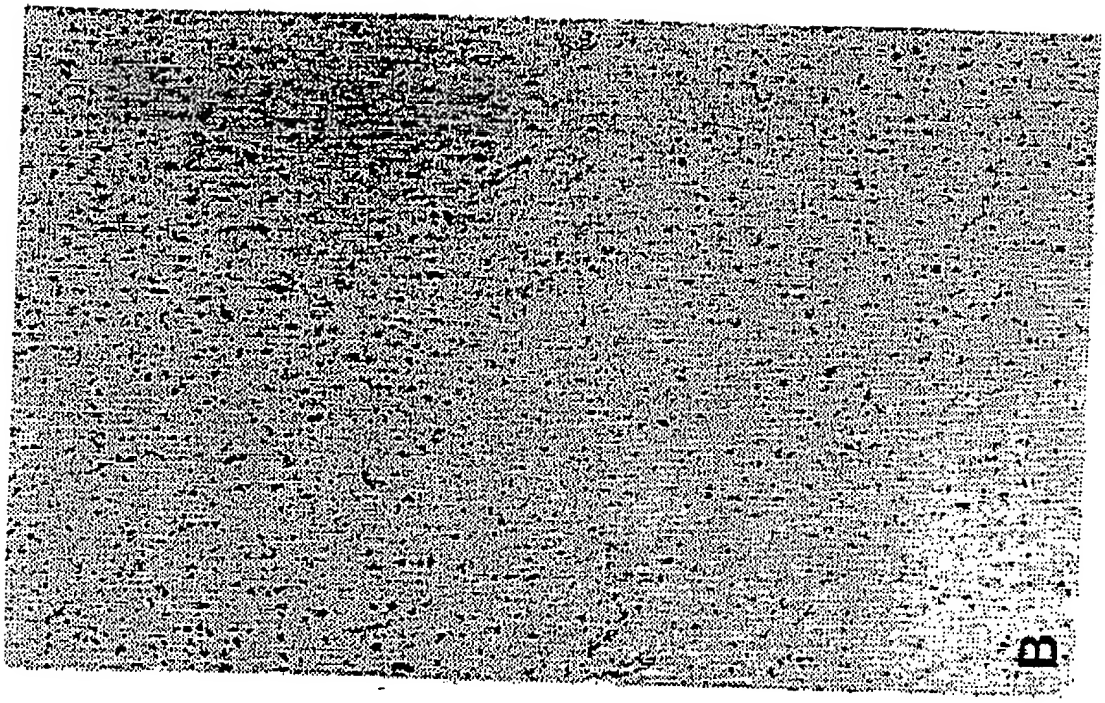


FIG. 6B

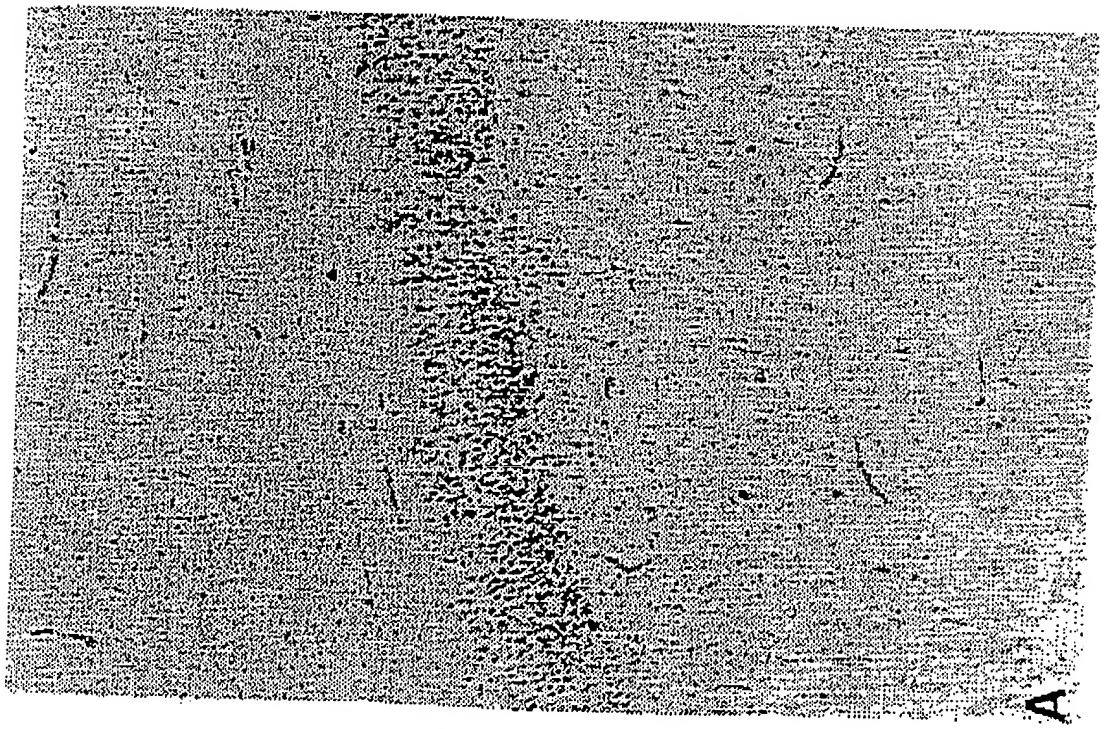


FIG. 6A

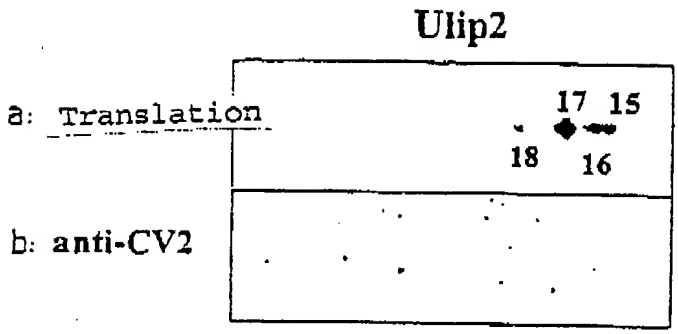


FIG. 7A

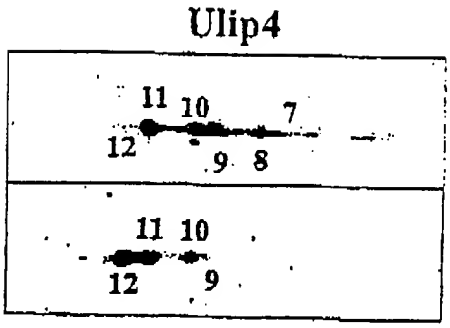


FIG. 7B

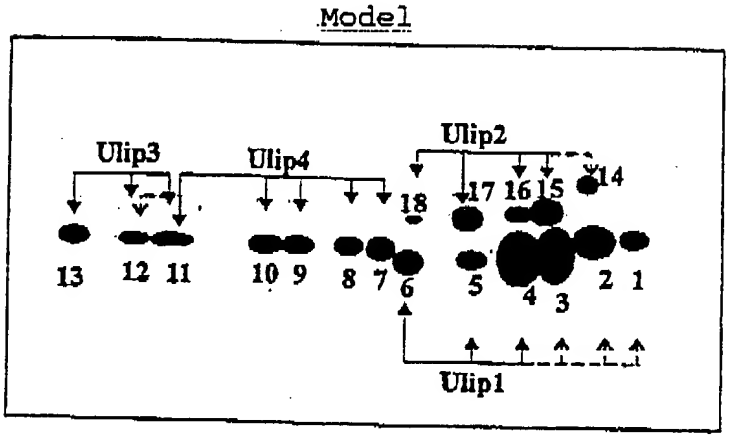


FIG. 7C

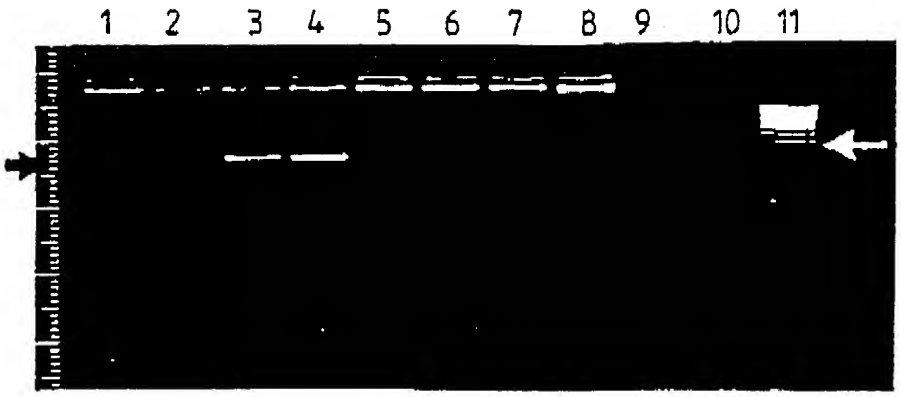


FIG. 8A

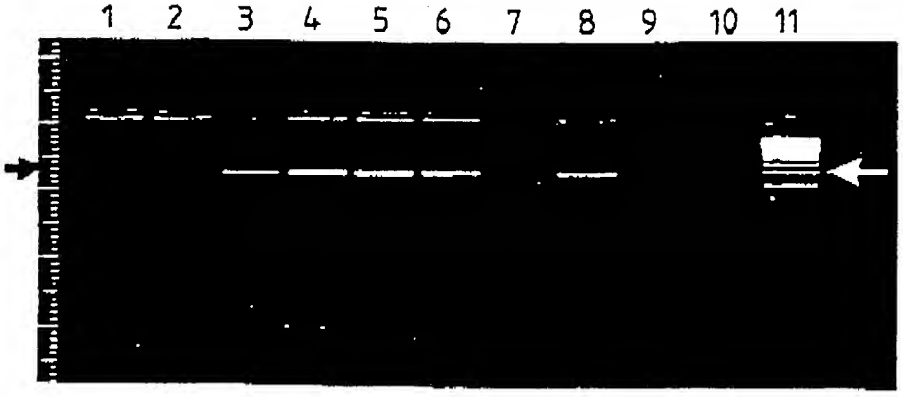


FIG.8B

ULIP2 mouse

```
                                  M   S   Y   Q   G   K   K   N   I   P   P   I      12
cttcctcccgccccccggagag ATG TCT TAT CAG GGG AAG AAA AAT ATT CCA CCC ATC              58

 T   S   D   R   L   L   I   K   G   G   K   I   V   N   D   D   Q   S      30
ACG AGC GAT CGT CTT CTG ATC AAA GGT GGC AAG ATT GTG AAT GAT GAC CAG TCC     112

 F   Y   A   D   I   Y   M   E   D   G   L   I   K   Q   I   G   E   N      48
TTC TAT GCA GAC ATA TAC ATG GAA GAT GGG TTG ATC AAG CAA ATA GGA GAA AAC     166

 L   I   V   P   G   G   V   K   T   I   E   A   H   S   R   M   V   I      66
CTG ATT GTA CCA GGA GGG GTG AAG ACC ATC GAA GCC CAC TCC AGA ATG GTG ATT     220

 P   G   G   I   D   V   H   T   R   F   Q   M   P   D   Q   G   M   T      84
CCC GGA GGA ATT GAC GTG CAT ACT CGC TTC CAG ATG CCT GAC CAG GGA ATG ACA     274

 S   A   D   D   F   F   Q   G   T   K   A   A   L   A   G   G   T   T     102
TCC GCT GAT GAC TTC TTC CAG GGA ACC AAG GCG GCC CTG GCC GGG GGA ACC ACC     328

 M   I   I   D   H   V   V   P   E   P   G   T   S   L   L   A   A   F     120
ATG ATC ATT GAC CAT GTT GTT CCT GAG CCC GGG ACG AGC CTA TTG GCT GCC TTT     382

 D   Q   W   R   E   W   A   D   S   K   S   C   C   D   Y   S   L   H     138
GAT CAG TGG AGG GAG TGG GCT GAC AGC AAG TCC TGC TGT GAC TAT TCG CTG CAC     436

 V   D   I   T   E   W   H   K   G   I   Q   E   E   M   E   A   L   V     156
GTG GAC ATC ACT GAG TGG CAC AAG GGC ATC CAG GAG GAG ATG GAA GCT CTG GTG     490
```

FIG. 9

K	D	H	G	V	N	S	F	L	V	Y	M	A	F	K	D	R	F	174
AAG	GAC	CAC	GGG	GTA	AAC	TCC	TTC	CTC	GTG	TAC	ATG	GCT	TTC	AAA	GAT	CGA	TTC	544
Q	L	T	D	S	Q	I	Y	E	V	L	S	V	I	R	D	I	G	192
CAG	CTG	ACG	GAT	TCC	CAG	ATC	TAT	GAA	GTG	CTG	AGC	GTG	ATC	CGG	GAT	ATC	GGT	598
A	I	A	Q	V	H	A	E	N	G	D	I	I	A	E	A	Q	Q	210
GCC	ATA	GCT	CAA	GTC	CAC	GCA	GAG	AAT	GGT	GAC	ATC	ATT	GCT	GAG	GCA	CAG	CAG	652
R	I	L	D	L	G	I	T	G	P	E	G	H	V	L	S	R	P	228
AGG	ATC	CTG	GAT	CTG	GGC	ATC	ACG	GGC	CCC	GAG	GGA	CAC	GTG	TTG	AGC	CGG	CCA	706
E	E	V	E	A	E	A	V	N	R	S	I	T	I	A	N	Q	T	246
GAG	GAG	GTC	GAG	GCT	GAA	GCT	GTG	AAC	CGG	TCC	ATC	ACT	ATT	GCC	AAC	CAG	ACC	760
N	C	P	L	Y	V	T	K	V	M	P	K	S	A	A	E	V	I	264
AAC	TGC	CCT	CTG	TAT	GTC	ACC	AAA	GTG	ATG	CCC	AAG	AGT	GCG	GCT	GAA	GTC	ATC	814
A	Q	A	R	K	K	G	T	V	V	Y	G	E	P	I	T	A	S	282
GCT	CAG	GCA	CGG	AAG	AAG	GGA	ACT	GTG	GTG	TAT	GGT	GAG	CCC	ATC	ACG	GCC	AGC	868
L	G	T	D	G	S	H	Y	W	S	K	N	W	A	K	A	A	A	300
CTG	GGG	ACT	GAT	GGC	TCT	CAT	TAC	TGG	AGC	AAG	AAC	TGG	GCC	AAG	GCT	GCG	GCC	922
F	V	T	S	P	P	L	S	P	D	P	T	T	P	D	F	L	N	318
TTT	GTC	ACC	TCC	CCA	CCC	TTG	AGC	CCC	GAC	CCA	ACC	ACT	CCA	GAC	TTT	CTC	AAC	976
S	L	L	S	C	G	D	L	Q	V	T	G	S	A	H	C	T	F	336
TCG	TTG	CTG	TCC	TGT	GGA	GAC	CTC	CAA	GTC	ACT	GGC	AGT	GCC	CAC	TGC	ACC	TTC	1030

FIG. 9 continued

```
 N   T   A   Q   K   A   V   G   K   D   N   F   T   L   I   P   E   G    354
AAC ACT GCC CAG AAG GCT GTG GGG AAG GAC AAC TTC ACC TTG ATT CCC GAG GGC   1084

 T   N   G   T   E   E   R   M   S   V   I   W   D   K   A   V   V   T    372
ACC AAC GGC ACT GAG GAG CGG ATG TCT GTC ATT TGG GAT AAA GCT GTG GTC ACT   1138

 G   K   M   D   E   N   Q   F   V   A   V   T   S   T   N   A   A   K    390
GGG AAG ATG GAT GAG AAT CAG TTT GTG GCT GTG ACC AGC ACC AAC GCA GCC AAA   1192

 V   F   N   L   Y   P   R   K   G   R   I   S   V   G   S   D   A   D    408
GTC TTC AAC CTT TAC CCC CGG AAA GGT CGC ATC TCG GTG GGA TCT GAT GCT GAC   1246

 L   V   I   W   D   P   D   S   V   K   T   I   S   A   K   T   H   N    426
TTG GTC ATC TGG GAC CCT GAC AGT GTG AAG ACC ATC TCT GCC AAG ACA CAC AAC   1300

 S   A   L   E   Y   N   I   F   E   G   M   E   C   R   G   S   P   L    444
AGT GCT CTT GAG TAC AAC ATC TTT GAA GGC ATG GAG TGT CGC GGC TCC CCA CTG   1354

 V   V   I   S   Q   G   K   I   V   L   E   D   G   T   L   H   V   T    462
GTG GTC ATC AGC CAG GGC AAG ATT GTC CTG GAG GAC GGC ACA CTT CAT GTC ACT   1408

 E   G   S   G   R   Y   I   P   R   K   P   F   P   D   F   V   Y   K    480
GAA GGC TCA GGA CGC TAC ATT CCC CGG AAG CCC TTC CCT GAC TTT GTG TAC AAA   1462

 R   I   K   A   R   S   R   L   A   E   L   R   G   V   P   R   G   L    498
CGC ATC AAA GCA AGG AGC AGG CTG GCT GAG CTG AGA GGG GTC CCT CGT GGC CTG   1516

 Y   D   G   P   V   C   E   V   S   V   T   P   K   T   V   T   P   A    516
TAT GAC GGA CCG GTA TGC GAG GTG TCT GTG ACG CCC AAG ACG GTG ACT CCA GCC   1570
```

FIG.9 continued

```
 S   S   A   K   T   S   P   A   K   Q   Q   A   P   P   V   R   N   L    534
TCA TCA GCT AAG ACA TCC CCT GCC AAG CAG CAG GCA CCA CCT GTT CGG AAC CTG  1624

 H   Q   S   G   F   S   L   S   G   A   Q   I   D   D   N   I   P   R    552
CAC CAG TCT GGA TTC AGC TTG TCT GGT GCT CAG ATT GAC GAC AAC ATT CCC CGC  1678

 R   T   T   Q   R   I   V   A   P   P   G   G   R   A   N   I   T   S    570
CGC ACC ACC CAG CGC ATC GTG GCA CCC CCT GGT GGC CGT GCC AAC ATC ACC AGC  1732

 L   G   *                                                               573
CTG GGC TAA agcccctaggcctgcaggccacttggggatgggggatgggacacctgaggacattctga 1800

gacttcctttcttccat
```

FIG. 9 continued

ULIP3 mouse

```
gctgtctgtcttcagcgccctcctctcgccctgcctctccctcctcctcccgccctccttgccaagccgggc    72

ggtgcaggcagccggagcagcggcggcgggccgagcagcggggagtgggcagcggtgggagccgagcttctg   144

tcctttctttcatccctccctggcctttgtcgccgctctcacgagtagcgccgccgggagagacccgggtag   216

                                                          M   S   H   Q       4
agcgccaggcagacgttagttccagcggccgggcggagggctccagaggggcc ATG TCT CAT CAG     281

 G   K   K   S   I   P   H   I   T   S   D   R   L   L   I   R   G   G      22
GGG AAG AAG AGC ATC CCG CAC ATC ACC AGT GAC CGG CTC CTC ATC AGA GGT GGA     335

 R   I   I   N   D   D   Q   S   F   Y   A   D   V   Y   L   E   D   G      40
CGC ATC ATC AAT GAT GAC CAG TCC TTC TAC GCC GAT GTC TAC CTA GAA GAT GGA     389

 L   I   K   Q   I   G   E   N   L   I   V   P   G   G   V   K   T   I      58
CTC ATA AAA CAA ATA GGA GAG AAC CTG ATT GTT CCT GGT GGA GTG AAG ACC ATC     443

 E   A   N   G   R   M   V   I   P   G   G   I   D   V   N   T   Y   L      76
GAG GCG AAT GGC CGA ATG GTC ATT CCC GGT GGC ATT GAT GTC AAC ACT TAC CTG     497

 Q   K   P   S   Q   G   M   T   S   A   D   D   F   F   Q   G   T   K      94
CAG AAG CCC TCC CAG GGC ATG ACC TCG GCT GAT GAC TTC TTC CAG GGC ACT AAA     551

 A   A   L   A   G   G   T   T   M   I   I   D   H   V   V   P   E   P     112
GCA GCG CTG GCA GGT GGA ACC ACG ATG ATC ATT GAC CAC GTT GTT CCT GAA CCT     605

 G   S   S   L   L   T   S   F   E   K   W   H   E   A   A   D   T   K     130
GGG TCC AGC TTG TTG ACT TCC TTT GAG AAA TGG CAC GAA GCA GCA GAC ACC AAA     659
```

FIG. 10

```
 S   C   C   D   Y   S   L   H   V   D   I   T   S   W   Y   D   G   V    148
TCC TGC TGT GAC TAT TCC CTC CAC GTG GAC ATC ACA AGC TGG TAT GAT GGT GTT    713

 R   E   E   L   E   V   L   V   Q   D   K   G   V   N   S   F   Q   V    166
CGG GAA GAG CTG GAG GTG CTG GTG CAG GAC AAA GGT GTC AAC TCC TTC CAA GTC    767

 Y   M   A   Y   K   D   L   Y   Q   M   S   D   S   Q   L   Y   E   A    184
TAC ATG GCG TAT AAG GAC CTG TAC CAG ATG TCT GAC AGC CAG CTG TAT GAA GCC    821

 F   T   F   L   K   G   L   G   A   V   I   L   V   H   A   E   N   G    202
TTC ACC TTC CTT AAG GGT TTG GGA GCT GTG ATC TTA GTC CAT GCA GAA AAT GGA    875

 D   L   I   A   Q   E   Q   K   R   I   L   E   M   G   I   T   G   P    220
GAT TTG ATA GCT CAG GAA CAA AAA CGG ATC CTG GAG ATG GGC ATC ACG GGT CCC    929

 E   G   H   A   L   S   R   P   E   E   L   E   A   E   A   V   F   R    238
GAG GGT CAT GCT CTG AGC AGA CCC GAG GAG CTG GAG GCC GAG GCT GTG TTC CGG    983

 A   I   A   I   A   G   R   I   N   C   P   V   Y   I   T   K   V   M    256
GCT ATT GCC ATT GCA GGC CGG ATC AAT TGC CCT GTG TAC ATC ACC AAG GTC ATG   1037

 S   K   S   A   A   D   I   I   A   L   A   R   K   K   G   P   L   V    274
AGC AAG AGT GCA GCG GAC ATC ATC GCA CTG GCC AGG AAG AAA GGC CCT CTT GTC   1091

 F   G   E   P   I   A   A   S   L   G   T   D   G   T   H   Y   W   S    292
TTC GGT GAG CCC ATA GCC GCC AGC CTG GGA ACC GAT GGC ACC CAC TAC TGG AGC   1145

 K   N   W   A   K   A   A   A   F   V   T   S   P   P   L   S   P   D    310
AAG AAC TGG GCC AAG GCA GCT GCA TTT GTG ACT TCC CCT CCC CTG AGC CCA GAC   1199
```

FIG.10 continued

```
 P   T   T   P   D   Y   L   T   S   L   L   A   C   G   D   L   Q   V    328
CCC ACC ACT CCT GAC TAC TTG ACC TCC TTG CTG GCC TGT GGA GAC TTG CAG GTC  1253

 T   G   S   G   H   C   P   Y   S   I   A   Q   K   A   V   G   K   D    346
ACA GGT AGT GGC CAC TGT CCC TAC AGT ATT GCT CAG AAG GCT GTG GGC AAG GAC  1307

 N   F   T   L   I   P   E   G   V   N   G   I   E   E   R   M   T   V    364
AAC TTC ACT CTG ATC CCT GAG GGT GTC AAT GGT ATA GAA GAG CGG ATG ACC GTT  1361

 V   W   D   K   A   V   A   T   G   K   M   D   E   N   Q   F   V   A    382
GTC TGG GAC AAG GCA GTG GCT ACT GGC AAG ATG GAT GAG AAC CAG TTT GTA GCC  1415

 V   T   S   T   N   A   A   K   I   F   N   L   Y   P   R   K   G   R    400
GTC ACC AGC ACC AAC GCA GCC AAG ATC TTC AAC CTG TAC CCG AGG AAA GGT CGG  1469

 I   A   V   G   S   D   A   D   V   V   I   W   D   P   D   K   M   K    418
ATC GCT GTG GGC TCC GAT GCT GAC GTA GTC ATC TGG GAC CCA GAT AAG ATG AAG  1523

 T   I   T   A   K   S   H   K   S   T   V   E   Y   N   I   F   E   G    436
ACC ATA ACA GCC AAA AGC CAT AAA TCA ACT GTG GAG TAC AAC ATC TTT GAG GGC  1577

 M   E   C   H   G   S   P   L   V   V   I   S   Q   G   K   I   V   F    454
ATG GAG TGC CAC GGC TCC CCC CTG GTG GTC ATC AGT CAG GGC AAG ATT GTC TTT  1631

 E   D   G   N   I   S   V   S   K   G   M   G   R   F   I   P   R   K    472
GAG GAT GGA AAC ATC AGT GTC AGC AAG GGC ATG GGC CGC TTC ATC CCT CGG AAG  1685

 P   F   P   E   H   L   Y   Q   R   V   R   I   R   S   K   V   F   G    490
CCA TTC CCA GAG CAT CTC TAC CAG CGT GTC AGG ATC AGA AGC AAG GTT TTC GGG  1739
```

FIG.10 continued

```
 L   H   S   V   S   R   G   M   Y   D   G   P   V   Y   E   V   P   A    508
TTG CAT AGT GTT TCC AGG GGC ATG TAC GAT GGG CCT GTG TAC GAG GTG CCA GCT  1793

 T   P   K   H   A   A   P   A   P   S   A   E   S   S   P   S   K   H    526
ACA CCC AAA CAT GCT GCT CCT GCT CCT TCT GCC GAA TCC TCG CCT TCT AAA CAC  1847

 Q   P   P   P   I   R   N   L   H   Q   S   N   F   S   L   S   G   A    544
CAA CCC CCA CCC ATC CGG AAC CTC CAC CAG TCC AAC TTC AGC TTA TCA GGT GCC  1901

 Q   I   D   D   N   N   P   R   R   T   G   H   R   I   V   A   P   P    562
CAG ATA GAT GAC AAC AAT CCA AGG CGT ACA GGC CAC CGC ATT GTG GCG CCC CCT  1955

 G   G   R   S   N   I   T   S   L   G   *                               573
GGT GGC CGC TCC AAC ATC ACC AGC CTC GGT TGA  cctcagatgagccagatatgcaagagt 2015

gaaggattatgggaaaacgtccattccttttccgtgtttttgaagcccacagttttagttggtactgacgga 2087

ggggagattgagcgatgctctttccttctctgtttaggaagaagtggtactagtgtggtgtgtttgcctgga 2159

agtccctcgcccacagtgtgtgttcacaccgactccacctcagagcatggtgccgtccgttttcccttcctа 2231

gtgaccccaggtttagcatcgtcctatactgttccctccactcctccatgaccctctgagtgatgg       2297
```

FIG.10 continued

ULIP4 | mouse

```
gctgactaatatgcttaaattcagcgggtcgccacgtctggtcggtacgtccacgcccgcgcagcccctacc    72

                                    M   S   F   Q   G   K   K   S   I   P      10
gaggacactcagcccgcccgtgtatcagg ATG TCC TTC CAA GGC AAG AAG AGC ATT CCC      131

 R   I   T   S   D   R   L   L   I   K   G   G   K   I   V   N   D   D      28
CGG ATA ACG AGC GAC CGC CTT CTC ATC AAA GGT GGG AAG ATT GTG AAC GAT GAC    185

 Q   S   F   H   A   D   L   Y   V   E   D   G   L   I   K   Q   I   G      46
CAG TCC TTT CAT GCT GAT CTG TAT GTG GAA GAC GGT CTG ATT AAA CAA ATT GGA    239

 E   N   L   I   V   P   G   G   I   K   T   I   D   A   H   G   L   M      64
GAA AAT CTC ATC GTC CCT GGG GGC ATC AAA ACC ATC GAT GCT CAT GGC CTG ATG    293

 V   L   P   G   G   V   D   V   H   T   R   L   Q   M   P   V   M   G      82
GTG CTG CCT GGG GGA GTT GAC GTT CAC ACC CGG CTG CAG ATG CCT GTG ATG GGC    347

 M   T   P   A   D   D   F   C   Q   G   T   K   A   A   L   A   G   G     100
ATG ACC CCA GCT GAT GAT TTC TGT CAG GGC ACC AAG GCG GCT CTA GCA GGC GGG    401

 T   T   M   I   L   D   H   V   F   P   D   A   G   V   S   L   L   A     118
ACC ACC ATG ATA TTG GAC CAT GTG TTT CCT GAC GCT GGT GTG AGC CTG CTG GCA    455

 A   Y   E   Q   W   R   D   G   A   D   S   A   A   C   C   D   Y   S     136
GCC TAT GAG CAG TGG CGG GAC GGA GCA GAC AGC GCG GCC TGC TGT GAC TAC TCC    509

 L   H   V   D   I   P   R   W   H   E   S   T   K   E   E   L   E   A     154
TTA CAT GTG GAC ATT CCT CGC TGG CAC GAG AGC ACC AAA GAA GAG CTG GAG GCC    563
```

FIG.11

```
 L   V   R   D   K   G   V   N   S   F   L   V   F   M   A   Y   K   D   172
CTA GTC AGG GAC AAA GGT GTG AAC TCC TTC CTG GTC TTC ATG GCA TAC AAG GAC   617

 R   C   Q   C   T   D   G   Q   I   Y   E   I   F   S   L   I   R   D   190
AGG TGC CAG TGT ACT GAC GGT CAG ATA TAT GAA ATC TTC AGC CTC ATC CGG GAC   671

 L   G   A   V   A   Q   V   H   A   E   N   G   D   I   V   E   E   E   208
CTG GGA GCT GTG GCC CAG GTG CAC GCA GAA AAT GGG GAC ATC GTG GAG GAG GAA   725

 Q   K   R   L   L   E   Q   G   I   T   G   P   E   G   H   V   L   S   226
CAG AAG CGC CTG CTG GAG CAA GGC ATC ACT GGT CCT GAG GGC CAT GTG CTC AGC   779

 H   P   E   E   V   E   A   E   A   V   Y   R   A   V   T   I   A   K   244
CAC CCA GAA GAG GTA GAG GCC GAG GCT GTG TAC AGA GCA GTC ACC ATT GCC AAG   833

 Q   A   N   C   P   L   Y   V   T   K   V   M   S   K   G   A   A   D   262
CAG GCC AAC TGC CCA CTA TAC GTC ACC AAG GTG ATG AGC AAG GGT GCA GCT GAC   887

 M   V   A   Q   A   K   R   R   G   V   V   V   F   G   E   P   I   T   280
ATG GTT GCC CAA GCC AAG CGC AGG GGG GTG GTC GTC TTT GGG GAA CCT ATC ACT   941

 A   S   L   G   T   D   G   S   H   Y   W   S   K   N   W   A   K   A   298
GCC AGC CTG GGC ACT GAT GGC TCA CAC TAC TGG AGC AAG AAC TGG GCC AAG GCT   995

 A   A   F   V   T   S   P   P   I   N   P   D   P   T   T   A   D   H   316
GCA GCC TTT GTC ACT TCA CCC CCT ATC AAC CCG GAC CCT ACT ACT GCA GAC CAC  1049

 L   T   S   L   L   S   S   G   D   L   Q   V   T   G   S   A   H   C   334
CTC ACC TCT CTG CTG TCC AGT GGG GAC CTC CAG GTG ACA GGC AGT GCC CAC TGC  1103
```

FIG. 11 continued

T	F	T	T	A	Q	K	A	V	G	K	D	N	F	T	L	I	P	352
ACC	TTC	ACT	ACT	GCC	CAG	AAG	GCT	GTT	GGC	AAA	GAC	AAC	TTC	ACA	CTG	ATC	CCC	1157
E	V	V	N	G	I	E	E	R	M	S	V	V	W	E	K	C	V	370
GAG	GTA	GTC	AAC	GGT	ATA	GAA	GAG	CGC	ATG	TCT	GTG	GTC	TGG	GAG	AAA	TGT	GTG	1211
A	S	G	K	M	D	E	N	E	F	V	A	V	T	S	T	N	A	388
GCT	TCA	GGG	AAA	ATG	GAC	GAG	AAT	GAG	TTC	GTT	GCC	GTG	ACC	AGC	ACA	AAT	GCT	1265
A	K	I	F	N	F	Y	P	R	K	G	R	V	A	V	G	S	D	406
GCC	AAA	ATC	TTC	AAT	TTT	TAC	CCC	AGG	AAG	GGG	CGT	GTG	GCC	GTG	GGC	TCT	GAT	1319
A	D	L	V	I	W	N	P	R	A	T	K	V	I	S	A	K	S	424
GCT	GAC	CTG	GTC	ATC	TGG	AAC	CCC	AGG	GCC	ACG	AAA	GTC	ATC	TCT	GCC	AAG	AGC	1373
H	N	L	N	V	E	Y	N	I	F	E	G	V	E	C	R	G	V	442
CAT	AAC	CTG	AAT	GTA	GAG	TAC	AAC	ATC	TTT	GAA	GGA	GTG	GAG	TGC	CGA	GGA	GTG	1427
P	T	V	V	I	S	Q	G	R	V	V	L	E	D	G	N	L	L	460
CCC	ACG	GTG	GTC	ATA	AGT	CAG	GGC	AGA	GTG	GTG	CTG	GAG	GAC	GGA	AAC	CTG	CTT	1481
V	T	P	G	A	G	R	F	I	P	R	K	T	F	P	D	F	V	478
GTC	ACT	CCA	GGG	GCT	GGC	CGC	TTC	ATT	CCC	CGG	AAG	ACG	TTC	CCG	GAC	TTT	GTC	1535
Y	K	R	I	K	A	R	N	R	L	A	E	I	H	G	V	P	R	496
TAT	AAG	AGG	ATA	AAG	GCT	CGC	AAC	AGG	CTA	GCA	GAG	ATC	CAC	GGT	GTG	CCT	CGA	1589

FIG.11 continued

```
 G   L   Y   D   G   P   V   H   E   V   M   L   P   A   K   P   G   S    514
GGC CTG TAC GAC GGG CCT GTG CAT GAA GTG ATG TTA CCT GCC AAG CCA GGA AGT  1643

 G   T   Q   A   R   A   S   C   S   G   K   I   S   V   P   P   V   R    532
GGC ACA CAG GCC CGT GCA TCC TGT TCA GGC AAG ATC TCA GTG CCA CCC GTG CGC  1697

 N   L   H   Q   S   G   F   S   L   S   G   S   Q   A   D   D   H   I    550
AAC CTG CAC CAG TCG GGG TTC AGC CTA TCT GGC TCT CAG GCT GAC GAT CAC ATT  1751

 A   R   R   T   A   Q   K   I   M   A   P   P   G   G   R   S   N   I    568
GCC AGA CGT ACG GCT CAG AAG ATC ATG GCA CCC CCC GGA GGA CGC TCC AAC ATC  1805

 T   S   L   S   *                                                       573
ACG TCT CTT TCC TAG  acttggggtcttggcaagctggtgctgtccccactggcagggtgtggggac 1871

gactcacgtcagttagctccttcctttgtagattgttattgtgaaaggc                       1920
```

FIG.11 continued

ULIP4 man

```
                                                  M   S   F   Q   G   K   K   S   I   P     10
GCCGCCCCTACCAGAGACCCCCAGGAGCAGG ATG TCC TTC CAG GGC AAG AAA AGC ATC CCC    61

 R   I   T   S   D   R   L   L   I   R   G   G   R   I   V   N   D   D     28
CGG ATC ACG AGT GAC CGC CTT CTG ATC AGA GGT GGG AGG ATC GTG AAT GAC GAC   115

 Q   S   F   Y   A   D   V   H   V   E   D   G   L   I   K   Q   I   G     46
CAG TCC TTT TAC GCT GAT GTG CAC GTG GAA GAT GGC TTG ATA AAA CAA ATC GGA   169

 E   N   L   I   V   P   G   G   I   *   T   I   D   A   H   G   L   M     64
GAA AAC CTC ATC GTC CCT GGG GGC ATC TAG ACC ATT GAC GCC CAC GGC CTG ATG   223

 V   L   P   G   G   V   D   V   H   T   R   L   Q   M   P   V   L   G     82
GTC CTT CCT GGT GGC GTT GAC GTC CAC ACA AGG CTG CAG ATG CCT GTC CTG GGC   277

 M   T   P   A   D   D   F   C   Q   G   T   K   A   A   L   A   G   G    100
ATG ACA CCG GCT GAC GAC TTC TGT CAG GGC ACC AAG GCA GCG CTA GCA GGA GGA   331

 T   T   M   I   L   D   H   V   F   P   D   T   G   V   S   L   L   A    118
ACC ACC ATG ATC TTG GAC CAC GTC TTC CCC GAC ACG GGT GTG AGC CTG CTG GCG   385

 A   Y   E   Q   W   R   E   R   A   D   S   A   A   C   C   D   Y   S    136
GCC TAC GAG CAG TGG CGG GAG CGG GCG GAC AGC GCG GCC TGC TGC GAC TAC TCC   439

 L   H   V   D   I   T   R   W   H   E   S   I   K   E   E   L   E   A    154
CTG CAC GTG GAC ATC ACC CGA TGG CAT GAG AGC ATC AAG GAG GAG CTG GAG GCC   493
```

FIG. 12

```
 L   V   K   E   K   G   V   N   S   F   L   V   F   M   A   Y   K   D     172
CTG GTC AAG GAG AAG GGT GTG AAC TCC TTC CTG GTC TTC ATG GCA TAC AAG GAC    547

 R   C   Q   C   S   D   S   Q   M   Y   E   I   F   S   I   I   R   D     190
CGG TGC CAG TGC AGC GAC AGC CAG ATG TAC GAG ATC TTC AGC ATC ATC CGG GAC    601

 L   G   A   L   A   Q   V   H   A   E   N   G   D   I   V   E   E   E     208
CTG GGG GCC TTG GCC CAG GTG CAC GCT GAG AAC GGG GAC ATC GTG GAG GAG GAG    655

 Q   K   R   L   L   E   L   G   I   T   G   P   E   G   H   V   L   S     226
CAG AAG CGG TTG CTG GAG CTC GGC ATC ACT GGC CCC GAG GGC CAC GTG CTC AGC    709

 H   P   E   E   V   E   A   E   A   V   Y   R   A   V   T   I   A   K     244
CAC CCC GAG GAG GTG GAG GCT GAG GCG GTG TAC CGA GCT GTC ACC ATC GCC AAG    763

 Q   A   N   C   P   L   Y   V   T   K   V   M   S   K   G   A   A   D     262
CAG GCA AAC TGC CCG CTG TAC GTC ACC AAG GTG ATG AGC AAG GGG GCG GCC GAC    817

 A   I   A   Q   A   K   R   R   G   V   V   V   F   G   E   P   I   T     280
GCC ATC GCT CAG GCC AAG CGC AGA GGG GTG GTC GTG TTT GGG GAG CCC ATC ACC    871

 A   S   L   G   T   D   G   S   H   Y   W   S   K   N   W   A   K   A     298
GCC AGC CTG GGC ACC GAC GGT TCA CAC TAC TGG AGC AAG AAC TGG GCC AAG GCT    925

 A   A   F   V   T   S   P   P   V   N   P   D   P   T   T   A   D   H     316
GCA GCC TTC GTC ACA TCA CCC CCT GTC AAC CCA GAC CCC ACC ACG GCA GAC CAC    979

 L   T   C   L   L   S   S   G   D   L   Q   V   T   G   S   A   H   C     334
CTC ACC TGC TTG CTG TCC AGC GGG GAC CTC CAG GTG ACA GGC AGC GCC CAC TGC   1033
```

FIG.12 continued

```
 T   F   T   T   A   Q   K   A   V   G   K   D   N   F   A   L   I   P   352
ACC TTC ACC ACT GCC CAG AAG GCT GTG GGC AAG GAC AAC TTC GCG CTG ATC CCC  1087

 E   G   T   N   G   I   E   E   R   M   S   M   V   W   E   K   C   V   370
GAG GGC ACC AAC GGC ATT GAG GAG CGC ATG TCG ATG GTC TGG GAG AAA TGT GTG  1141

 A   S   G   K   M   D   E   N   E   F   V   A   V   T   S   T   N   A   388
GCC TCT GGG AAG ATG GAC GAG AAT GAG TTC GTC GCG GTG ACC AGT ACA AAT GCT  1195

 A   K   I   F   N   F   Y   P   R   K   G   R   V   A   V   G   S   D   406
GCC AAA ATC TTC AAT TTT TAC CCA AGG AAG GGG CGA GTG GCT GTG GGC TCT GAC  1249

 A   D   L   V   I   W   N   P   K   A   T   K   I   I   S   A   K   T   424
GCT GAC CTG GTC ATA TGG AAC CCC AAG GCC ACC AAG ATC ATC TCT GCC AAG ACC  1303

 H   N   L   N   V   E   Y   N   I   F   E   G   V   E   C   R   G   A   442
CAC AAT CTG AAC GTG GAG TAC AAC ATC TTC GAG GGA GTG GAG TGC CGG GGA GCG  1357

 P   A   V   V   I   S   Q   G   R   V   A   L   E   D   G   K   M   F   460
CCT GCC GTG GTC ATA AGT CAG GGC CGA GTG GCG CTG GAG GAC GGG AAG ATG TTT  1411

 V   T   P   G   A   G   R   F   V   P   R   K   T   F   P   D   F   V   478
GTC ACC CCG GGG GCG GGC CGC TTC GTC CCT CGG AAA ACA TTC CCG GAC TTT GTC  1465

 Y   K   R   I   K   A   R   N   R   L   A   E   I   H   G   V   P   R   496
TAC AAG AGG ATC AAA GCT CGC AAC AGG CTG GCG GAG ATC CAC GGT GTG CCC CGT  1519
```

FIG.12 continued

```
 G   L   Y   D   G   P   V   H   E   V   M   V   P   A   K   P   G   S    514
GGG CTG TAT GAC GGG CCC GTC CAC GAG GTG ATG GTG CCT GCC AAG CCA GGG AGT   1573

 G   A   P   A   R   A   S   C   P   G   K   I   S   V   P   P   V   R    532
GGC GCT CCG GCC CGC GCG TCC TGC CCA GGC AAG ATC TCC GTG CCT CCT GTG CGC   1627

 N   L   H   Q   S   G   F   S   L   S   G   S   Q   A   D   D   H   I    550
AAC CTA CAT CAG TCG GGG TTC AGC CTA TCT GGG TCT CAG GCT GAT GAC CAC ATC   1681

 A   R   R                                                               553
GCC CGA CGC                                                             1690
```

FIG.12 continued